REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
7 May 1998

3. REPORT TYPE AND DATES COVERED

4. TITLE AND SUBTITLE
ERROR TYPES AND RELATED ERROR DETECTION PROCESSES IN THE AVIATION DOMAIN

5. FUNDING NUMBERS

6. AUTHOR(S)
Heather M. Alexander

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

8. PERFORMING ORGANIZATION REPORT NUMBER
98-008

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
THE DEPARTMENT OF THE AIR FORCE
AFIT/CIA, BLDG 125
2950 P STREET
WPAFB OH 45433

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION AVAILABILITY STATEMENT
Unlimited distribution
In Accordance With AFI 35-205/AFIT Sup 1

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

19980514 000

14. SUBJECT TERMS

15. NUMBER OF PAGES
68

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT

18. SECURITY CLASSIFICATION OF THIS PAGE

19. SECURITY CLASSIFICATION OF ABSTRACT

20. LIMITATION OF ABSTRACT

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

THE GRADUATE COLLEGE

MAY 1998

WE HEREBY RECOMMEND THAT THE THESIS BY

HEATHER M. ALEXANDER

ENTITLED ERROR TYPES AND RELATED ERROR DETECTION PROCESSES IN THE AVIATION DOMAIN

BE ACCEPTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR

THE DEGREE OF MASTER OF SCIENCE

Nadine Sarter

Director of Thesis Research

Head of Department

Committee on Final Examination†

Nadine Sarter Chairperson

Christopher D. Wickens

† Required for doctor's degree but not for master's.

O-517

ERROR TYPES AND RELATED ERROR DETECTION PROCESSES IN THE AVIATION DOMAIN

BY

HEATHER M. ALEXANDER

B.S., Norwich University, 1992

B.S., Norwich University, 1992

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Psychology
in the Graduate College of the
University of Illinois at Urbana-Champaign, 1998

Urbana, Illinois

Abstract

HEATHER M. ALEXANDER
Captain, United States Air Force, Biomedical Sciences Corps

ERROR TYPES AND RELATED ERROR DETECTION PROCESSES IN THE AVIATION DOMAIN

1998

Number of pages: 68 (including Bibliography)

Degree: Master of Science in Psychology

University of Illinois at Urbana-Champaign

Human error has been identified as a contributing factor in 75-80% of all aviation accidents. To date, most efforts to improve flight safety have focused on error prevention. A different approach that has received less attention is to avoid the negative consequences of erroneous actions and assessments by supporting their timely detection. In this study, aviation incidents were analyzed in terms of the type of error involved (errors of omission and commission; slips, lapses, and mistakes), the performance level at which the error occurred (skill-, rule-, or knowledge-based performance), and the relation between error types and error detection processes that prevented these incidents from turning into accidents. The majority of reported errors were lapses, i.e., failures to perform a required action, and mistakes, i.e., errors in the formation of an intention. Relatively few slips, i.e., inappropriate executions of intended actions, were reported. Slips appear to be detected and corrected by the pilot before they result in an unsafe situation that is worth reporting. Lapses and mistakes, on the other hand, are more difficult for the pilot committing the error to detect and, in most cases, required intervention by air traffic control. A large percentage of lapses resulted from inattention, either due to some distraction in the cockpit or due to multiple competing demands. Mistakes, on the other hand, frequently occurred as a consequence of some misunderstanding between pilots and air traffic controllers concerning clearances and intentions. Most lapses were detected incidentally based on routine checks of aircraft settings and performance, whereas errors of commission, which include both mistakes and slips, were detected equally often based on monitoring for the immediate outcome of an action and by routine checks. These findings indicate the need for more effective support of error detection, particularly in the case of lapses and mistakes. This goal may be achieved through enhanced feedback that captures the pilot's attention in case of a mismatch between intention and action, through improved air-ground coordination in the interest of shared knowledge of intent, or through procedures that minimize the potential for distractions on the flight deck.

Table of Contents

Chapter

List of Tables

List of Figures

Introduction

There is widespread concern in the aviation industry that the expected growth in air traffic during the next decade may lead to an average of one major airplane accident per week unless the current, already low accident rate is reduced even further (Schiavo, 1998). This dire prediction calls for all parties in the aviation industry as well as researchers in the area of human factors to explore new options for improving safety. Given that 75-80 % of all aviation incidents and accidents are viewed as the result of pilot error, one promising approach appears to be investments in a better understanding of the reasons for and possible countermeasures to erroneous actions and assessments.

There are three different ways to reduce the frequency or consequences of error: error prevention, error detection, and error tolerance. The prevention of errors through improved training, design, and procedures has been the focus of research and development activities for a long time. One example of an error prevention mechanism is limiting functions that do not allow an undesirable or unsafe action to occur or continue. It is important to keep in mind, however, that it is not possible to eliminate, or prevent, completely the occurrence of errors. Therefore, additional steps need to be taken that support operators in detecting and recovering from an error if and when it occurs.

Supporting error recovery requires systems that are error-tolerant, i.e., that immediate and irreversible negative consequences are avoided by allowing the operator to modify his or her initial input or action. For example, some word-processing systems are error-tolerant as they create a temporary back-up copy of a file that the operator can access if he/she inadvertently deleted the file. Error tolerance requires that the operator detects that an error was made in the first place.

In most cases, error detection is based on the realization that (the outcome of) an action is different from the intended or expected one. An operator may become aware of an error through a wide range of mechanisms and sources. Another individual may point out the error, or the operator may detect the (undesired outcome of an) action him/herself based on a check of the progress towards a goal. Checks may occur as part of routine behavior or due to a suspicion that something may not be correct. Something may

remind the operator of an action he/she forgot to perform, or a system may alert the operator to a problem through an alarm or message. Once an operator detects that an error has occurred, he/she can begin the process of identifying and trying to correct the error.

To date, the research and literature on human error has focused primarily on the identification of factors contributing to human error and on the development of different error classification schemes (Reason, 1990). Little empirical data exist on (the effectiveness of) error detection mechanisms and their relation to various error types and performance levels. Numerous questions deserve further investigation, including:

- What is the relationship between error types and error detection processes?
- What are the factors that cause detection failure, and how can error detection be enhanced?
- What forms the basis of the reference mechanisms against which actions or their consequences are checked?
- How does self-detection differ from detection of errors by other people?
- What are the group dynamics of error detection for real, complex systems where knowledge is distributed?

The goal of this thesis is to provide insight into the above questions based on an analysis of incident reports from the aviation domain. This domain was selected because it represents a rich source of information on errors in a highly complex, event-driven real-world environment. In aviation, many competing cognitive demands are placed on various operators (e.g., pilots, air traffic controllers) who need to coordinate their activities in the interest of safe and efficient flight operations. These individuals are highly constrained by procedures and regulations to help avoid errors that can have disastrous consequences for a large number of people. Still, errors occur and sometimes go unnoticed leading to incidents and accidents.

To examine the above questions we will analyze incident reports from the Aviation Safety Reporting System (ASRS). This methodological approach was chosen because ours is one of the first studies concerning error detection in the real-world

environment. For that reason, we are interested in exploring the entire range of naturally occurring error types and detection mechanisms. Findings from this work can provide input and guidance for future more controlled studies of error detection in simulated environments. In these controlled environments scenarios are designed to investigate specific errors and confirm predictions from earlier analysis and exploratory empirical work.

This study may also confirm findings from earlier research on error detection which has been conducted, for the most part, in the context of very specific, isolated, self-paced tasks that were performed by individuals in much simpler and less risky domains. It is not clear that the results of this earlier work transfer to environments such as aviation where the nature of errors and error detection mechanisms may be dissimilar due to different demands and constraints.

This thesis will address a number of questions related to human error. First, the nature and frequency of errors involved in the reported incidents will be examined. The phenotype or surface appearance of these errors and their outcome in aviation-specific terms will be described. Errors will be analyzed in terms of domain – independent characteristics and the cognitive stage at which they occur – slips at the level of execution, lapses related to breakdowns in storage, and mistakes involving errors in intention formation. Errors will also be classified as errors of omission which are equivalent to lapses and errors of commission which include slips and mistakes. This will allow us to compare the usefulness of different error categorization schemes for understanding, predicting and supporting error detection. Finally, errors will be analyzed in terms of the level of task performance at which they occur – skill-, rule-, or knowledge-based performance. The role and frequency of possible contributing factors to errors such as time pressure, distractions, or a lack of system understanding will be explored. Next, this research will examine the relationship between the different error types and the processes leading to their detection before the reported incident could turn into an accident. Questions addressed in this context are who is detecting the error and what mechanisms/information appear most effective for detecting the various error types.

Finally, this research will examine whether the ongoing introduction of increasingly advanced automated systems to the aviation domain has an impact on the nature and frequency of errors and on related error detection mechanisms.

A number of predictions concerning the above issues can be made based on earlier research. For example, it is anticipated that skill-based errors and commission errors which, includes slips, are detected rapidly and effectively by the operator committing the error since they are more likely to result in an observable outcome or feedback that is different from the well-defined outcome the operator is anticipating and monitoring for (Reason, 1990; Sellen, 1990). In contrast, the detection of errors of omission and of problem solving errors, or mistakes, is expected to require external intervention. In most cases, errors of omission fail to produce observable changes that can be compared to intentions. And, in the case of problem-solving tasks, the expected outcome is not always well-defined. The following sections will discuss in more detail the above mentioned error classification schemes, possible error detection mechanisms, and the hypotheses guiding this research.

Error Types and Error Detection Mechanisms

The Phenotype and Genotype of Error

Accidents and incidents are often investigated and reported in terms of their phenotype, i.e., in terms of their surface features and manifestation in domain-specific language (e.g., controlled flight into terrain, altitude deviations, or runway incursions). Classifying errors based on their surface appearance provides insight into the frequency of certain outcomes but it fails to identify common underlying mechanisms and therefore runs the risk of suggesting an unwieldy number of error categories (Hollnagel, 1993).

In order to understand and be able to mitigate (the effects of) errors, it is important to identify deeper and more general characteristics of observed difficulties – the genotype of error. Identifying lawful factors that shape the likelihood and nature of errors and their detection is a prerequisite for being able to predict, prevent, and manage them. In this research, we will therefore focus on the analysis of errors in terms of domain-independent categories that are related to their underlying cognitive mechanisms or the associated performance level.

Violations and Errors

One important distinction between different kinds of unsafe acts is the one between violations and errors. Violations are the "deliberate deviation of actions from safe operating procedures" (Reason, 1995). In other words, the act of committing a violation is intentional and performed for what appears to be a justifiable and necessary reason at the time. Violations tend to occur in a social context involving specific motivational factors such as organizational pressures.

Errors, on the other hand, are unintended actions. They are most often related to breakdowns in information processing rather than driven by motivational factors. A number of definitions of the term error have been proposed. For example, human error can be considered, "…a specific variety of human performance that is so clearly and significantly substandard and flawed when viewed in retrospect that there is no doubt that

it should have been viewed by the practitioner as substandard at the time the act was committed or omitted" (Woods et al, 1994). Reason (1990) defines human error as, "a planned sequence of mental or physical actions that fail to achieve the intended outcome, and when failure cannot be attributed to intervention by some chance agency". Reason integrates many of the different attempts to define human error by stating that "human error covers a wide variety of aberrant behaviors, where each type involves different psychological mechanisms, features in different parts of the system and demands different measures" (Reason, 1995).

Since the focus of this research is error detection, violations will not be included in this study. They are deliberate actions that may be inappropriate but do not require detection support. Our analysis will focus on errors exclusively.

Active and Latent Errors

Reason has introduced an important distinction between two different types of errors -- active and latent errors (Reason, 1990; Maurino, 1995). Latent errors or failures result from actions by people at the "blunt end" of a system (such as designers or managers) and may lie dormant in a system for an extended period of time until they combine with other factors to breach the system's defenses and create a problem. The detection of latent failures is primarily the goal and responsibility of software validation and system certification.

Active errors, on the other hand, involve some action or assessment by an operator "at the sharp end" (e.g., pilots, controllers). The effects of the errors tend to be felt almost immediately. These errors will be the focus of our research since their consequences can be mitigated by supporting operators in early, effective error detection.

Errors of omission and commission. Active errors can take the form of errors of omission and errors of commission. Omission errors are characterized by the failure to perform some required action. An operator may omit a step in a task, or omit the entire task. Commission errors, on the other hand, involve an operator who performs an action, but performs it in an inappropriate manner or at an inappropriate time. Commission errors

can take a wide variety of forms, including selection errors, sequence errors, and time errors. Selection errors occur when the operator chooses the wrong or inappropriate mechanism or tool for execution of the task. Sequence errors are errors in the order of execution of the individual actions required to attain the task goal. And time errors are errors in the time planned, or allotted, for completion of the task.

The omission-commission distinction is domain-independent but still does not take into account underlying or contributing psychological mechanisms. It is relevant in the context of this research since it has implications for the likelihood and form of error detection. In general, errors of commission are considered to be easier and faster to detect by the operator him/herself based on progress checks following an action. In contrast, errors of omission may go undetected since, in the absence of an action, monitoring for any changes or effects is not likely to occur. In most cases, the detection of an error of omission is expected to require an external source or agent.

Slips, lapses, and mistakes. Norman (1981) and Reason (1990) have proposed another, partially overlapping, classification of active errors. They distinguish between slips, lapses, and mistakes. Mistakes are errors in the formation of an intention or the choice of a method for achieving a goal, and are related to a breakdown in the planning stage. Slips and lapses, on the other hand, are errors in the execution of an intention. Slips occur when an intention is executed in an inappropriate manner whereas lapses represent the failure to perform some required action. Slips occur when there is a breakdown in the execution stage, while lapses are related to breakdowns in the storage stage.

Slips can be broken down further into description errors; actuation or triggering errors; and capture errors. Description errors result from the operator working at a level of abstraction that is higher than necessary for the task at hand. As an example, a slip can result in confusion of one control knob for another. Actuation or triggering errors are a failure of the operator to appropriately activate a necessary action including unintended activation, or loss of activation, of a schema. An example is failing to shift task goals from a primary task to a critical secondary task in a timely manner, or correctly timing the action but performing it in a reversed manner. Capture errors result from faulty

triggering of active schema, often as a result of habit intrusions. An example of a capture error is a pilot who is use to flying with a flight engineer on board and who turns around to communicate or delegate a task to the flight engineer when, in fact, his/her current aircraft does not have a flight engineer (Thompson, 1980; Norman, 1981).

Mistakes are deficiencies or failures in the judgmental and/or inferential processes involved in the formation of a plan or intention. They can also involve failures in the specification of the process or method by which to achieve the intended outcome. It is not relevant for the determination of a mistake whether the actions undertaken by the operator are appropriate and successful.

Note that there is an overlap between the first error classification - errors of omission and errors of commission - and the latter distinction between slips, lapses, and mistakes. Slips and mistakes can be considered errors of commission whereas lapses are equivalent to errors of omission. The distinction between slips and mistakes provides a more detailed account of the mechanisms underlying the different types of error of commission - an error in intention formation vs. an error in the execution of an intention. Detection of these two different kinds of errors of commission is likely to occur via different mechanisms and may not be equally likely. The following figure provides an overview of the relation between the different error forms that we have discussed up to this point, and the performance levels to be discussed next.

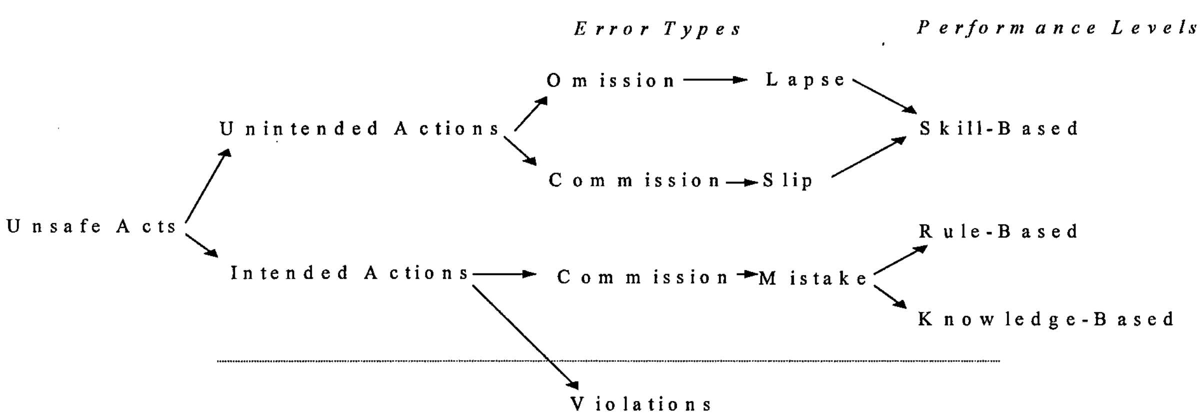

Figure 1: Depiction of Unsafe Acts (adapted from Reason, 1990)

Errors At Different Performance Levels

Note that figure 1 lays out not only different kinds of errors but also illustrates their relation to three different levels of performance which were first introduced by Rasmussen and Jensen (1974). Based on their research on supervisory control in industrial installations, Rasmussen differentiates between skill-, rule-, and knowledge-based performance.

Skill-based performance takes place in the case of highly practiced routine actions. These actions are carried out in an automatic fashion and are easy for the experienced operator to perform. The actions tend to occur rapidly and without intentional effort. Stored action and perception patterns which are acquired through training and experience are driving performance with errors occurring as a result of variability of force, space, or time coordination.

Rule-based behavior requires more conscious effort. It involves the application of stored solutions to familiar problems. These solutions take the form of "if A (state), then B (diagnosis/remedial action)". Errors at the rule-based level are typically associated with a misclassification of the situation, which then results in the misapplication of a good rule or the correct application of an inadequate rule due to the incorrect recall of procedures.

Performance at the knowledge-based level requires the greatest amount of conscious cognitive effort. This effort is directed at solving a novel problem for which no stored rules or procedures exist. Instead, a solution must be worked out by the operator(s) on-line. Errors at the knowledge-based performance level are the result of cognitive resource limitations and incomplete or incorrect knowledge of the situation (Reason, 1990).

Building on Rasmussen's work, Reason (1990) related the three earlier described error types – slips, lapses, and mistakes -- to these three performance levels. Slips and lapses occur at the skill-based performance level while mistakes are associated with either rule- or knowledge-based behavior. Skill-based errors, such as slips and lapses, occur while the operator is engaged in routine activities while rule- and knowledge-based mistakes take place once an operator engages in problem-solving behavior.

The likelihood of error detection is different for performance at the three different levels. Reason (1990) predicts that errors at the skill-based level (slips and lapses) are detected rapidly and effectively while the detection of rule- and knowledge-based mistakes is more difficult and often only achieved through intervention by some external source or agent.

Error Detection Sources and Mechanisms

Defenses in Depth. For complex high-risk systems, one important design principle is to implement several layers of protective mechanisms to ensure that negative consequences of errors are avoided, prevented, or deflected. This principle is referred to as "defenses-in-depth". It implies that several independent events have to coincide and several layers of system protections have to be penetrated before an error can result in an accident with disastrous consequences. In the context of this research, we can think of these layers as representing different error detection sources and mechanisms.

On the commercial flight deck, the first layer of defense is the pilot who can detect an error based on his/her own actions and their outcome or through various required checks. Should the pilot fail to detect the error, there is at least one other crewmember in the cockpit who may detect the error. If neither crewmember detects the error, the crew on another aircraft may notice and point out or address the problem. And finally, other remotely located operators may come into play. These operators include ground personnel such as dispatchers or air traffic controllers (see Figure 2).

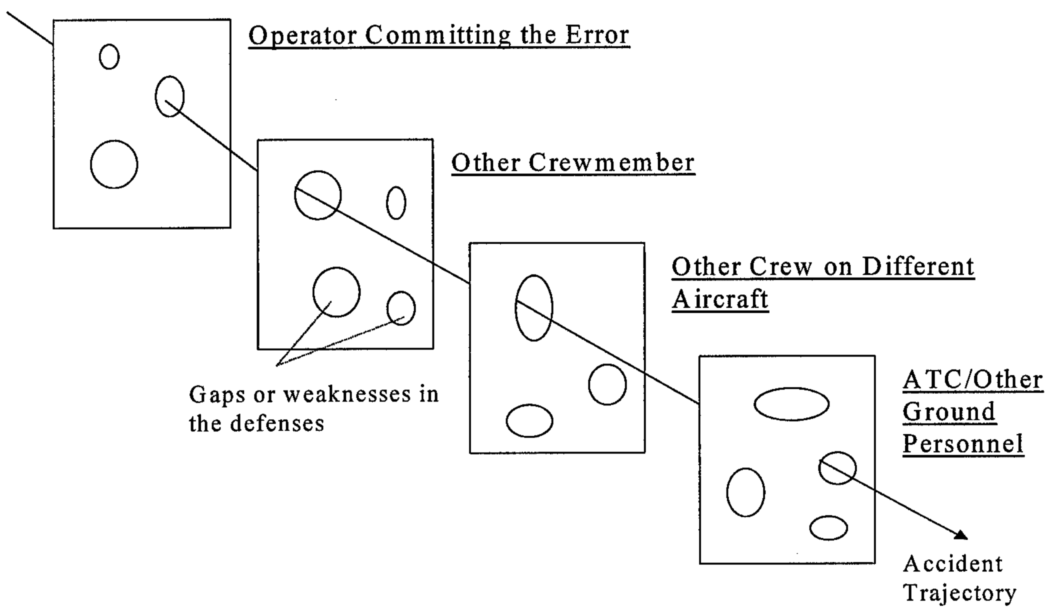

Figure 2: Defenses-in-Depth (adapted from Maurino, 1995)

So far, we have discussed possible sources of error detection – who detected the error?. It is also important to examine how or based on what information an error was noticed. Error detection can be based on knowledge-based information search or on data-driven attention capture. In the former case, an operator will search for information to confirm that the intended (outcome of an) action was indeed achieved. Error detection by the operator him/herself can occur prior to the execution of the erroneous action based on self-monitoring, i.e., based on the detection of a mismatch between stored representations of errors and the anticipated performance. This process occurs only in the context of well-known tasks and circumstances. In other words, it takes place in the context of skill-based performance (Sellen, 1990).

Error detection at a later stage occurs on the basis of feedback from overt actions. It involves three basic components: a) feedback regarding the actions or outcome of actions; b) an anticipated result or reference value for comparison; and c) a monitoring system that compares the feedback to the reference.

Error detection can also occur as a result of some other agent or some salient information in the operator's environment capturing his/her attention and alerting the operator to the problem. This alerting mechanism can be as simple as an alarm signal or a forcing function, or it can involve a highly intelligent system that attempts to infer pilot intent and compare it against the pilot's actions or input to determine if those actions/input are appropriate (Wickens, 1993). Data-driven attention capture may become increasingly important with the introduction of more and more independent and coupled automation which can take (possibly unanticipated) actions on its own (Sarter and Woods, 1995; Sarter, Woods, and Billings, 1997). To date, very little empirical research exists on the effectiveness of different error detection mechanisms and on their relation to different error types.

<u>Overview of Empirical Research On Error Types and Detection.</u> Most research on error detection has been conducted in laboratory settings and has looked at specific tasks such as typing (Rabbitt, 1978); reading comprehension (Kroll and Ford, 1992); go/no-go tasks with event-related brain potentials (Scheffers, et al., 1996); partial response (Coles, et al., 1995); speech (MacKay, 1992); choice-response (Rabbitt and Phillips, 1967); statistical problem solving (Allwood, 1984); visual search (Rabbitt, et al., 1978); and use of a computer database (Rizzo et al., 1987). The small number of studies that were performed in the context of real-world operational environments include nuclear power (Woods et al., 1987); maritime (Van Eekhout and Rouse, 1981); aviation (O'Hare et al., 1994; Wiegmann and Shappell, 1997; Degani et al., 1991); and hospitals (Barker, 1962).

Some of the above studies provide insight into some aspect of error detection, but their focus tends to be different from our areas of interest. For example, the study conducted by Scheffers examined event-related brain potentials in the context of errors made in a choice reaction task. Coles' research focused on the relation between the force of the response output and the level of uncertainty regarding that response. And Rabbitt's work examined the relation between the speed and accuracy of error correction. These studies investigated error detection and correction in a controlled environment that was

designed to limit the range and expression of possible errors and error detection mechanisms. Tasks were mostly self-paced and relatively simple with no, or few, competing task demands. In that sense, findings from those studies are of limited relevance to our objectives.

Research by Allwood (1984) provided the basis for many later studies on error detection and classification. The focus of Allwood's study was self-detection of errors. Sixteen subjects were instructed to think aloud while solving two statistical problems. Allwood found that, overall, 69% of the errors were detected. These errors were categorized as execution errors (62%), solution method errors (21%), skip errors (9%), higher level mathematical errors (5%), and other errors (2%). The execution errors in this study are equivalent to slips, solution method errors to rule-based errors, and higher level mathematical errors to knowledge-based mistakes.

The five categories of errors that were observed in this study were detected by means of three different mechanisms. Direct Error Hypothesis (DEH) behavior occurred when subjects suddenly detected a real or suspected error. This behavior did not always occur immediately following the error but could occur later in the process. Error Suspicious (ES) behavior took place when the subject noticed some result that was strange or unexpected. Some property or outcome was questioned without directly identifying it as an error. The third behavior was Standard Check behavior (SC) and was initiated by the subject independent of observing any suspicious outcome.

Overall, Allwood found that subjects in his study detected 87% of the execution errors (which are similar to Reason's slips) and 52% of the solution method errors (which are equivalent to mistakes). Direct error hypothesis behavior led to the detection of 64% of the execution errors and 23% of the solution method errors. Error suspicious evaluation behavior was involved in the detection of 22% of the execution errors and 26% of the solution method errors. A standard check led to detection of 2% of the execution errors and 5% of the solution method errors.

Based on the results of this study, Allwood argues for the existence of two basic types of error detection - a sudden direct detection method and detection by means of more elaborate processes. The sudden direct method corresponds to the standard check

(SC) behavior. The more elaborate method is assumed to be the result of an observed mismatch between actual and expected result of an action. The elaborate method involves detection arising from a DEH episode, i.e., action-based detection, or detection based on an ES episode, i.e., outcome-based error detection. In summary, the two major detection categories identified in this study are either self-detection or outcome-based detection. Allwood's study also suggests that omission errors are difficult to detect for the individual committing the error.

Another group of investigators further explored error detection using Allwood's error detection categories. Rizzo, et al. (1987) investigated the relationship between error types and patterns of error detection. The primary focus of their work was the cognitive processes underlying self-detection of errors. The study categorized observed behaviors and errors according to the skill-, rule-, and knowledge-based performance levels described by Rasmussen and Jensen (1974) and according to the error types proposed by Norman (1981) - slips, lapses, and mistakes. These were related to the error detection behavior patterns proposed by Allwood - Direct Error Hypothesis (DEH), Error Suspicious (ES), and Standard Check (SC).

Sixteen subjects were asked to think aloud while solving problems using a database system under two experimental conditions. During the course of the experiment the subjects were expected to detect, locate, and combine items in the database. In the first condition, the level of task complexity was varied over four experimental sessions. These sessions required consistent use of database manipulations that were expected to become automatic with increasing experience on the task. This shift towards skill-based behavior, was expected to result in fewer knowledge-based errors but more slips in the later sessions. In the second condition, the subject was required to find an item in the database. The item changed in each of the four sessions, but the record containing the item remained the same. Both conditions were designed to test the effect of practice as a function of attention allocation and the relationship between error types, patterns of error detection, and psychological mechanisms of detection.

Overall, the subjects in this study made 1277 errors and detected 1097 of those errors. The authors found that the subjects' error detection performance improved with

practice as their ability to use feedback and to specify intentions increased. The following tables summarize observed error types and associated detection mechanisms in this study.

	Slips/Lapses	Mistakes – Rule	Mistakes – Knowledge
Direct Error Hypothesis	72%	42%	15%
Standard Check	7%	9%	10%
Error Suspicious	4%	35%	57%
Undetected	16%	14%	18%

Table 1: Condition I Results

	Slips/Lapses	Mistakes – Rule	Mistakes – Knowledge
Direct Error Hypothesis	82%	82%	29%
Standard Check	5%	3%	2%
Error Suspicious	4%	10%	37%
Undetected	5%	6%	32%

Table 2: Condition II Results

The largest number of errors in both conditions were slips which were most often detected by means of DEH behavior. A large number of slips went undetected which Rizzo et al. explained by the distance between the level of specification of intention and the level of execution of the action. When an individual is performing at a knowledge-based level, large portions of attentional resources are directed to plan execution. This attentional demand makes the action prone to slips. The likelihood of detection of slips is considered the result of a trade-off between available resources and distance between levels of specification and action (Rizzo et al., 1987). Most rule-based mistakes were detected by means of DEH and ES behavior.

The detection of knowledge-based mistakes was particularly difficult for subjects in this study. For the most part, they were detected based on ES behavior. 18% of the knowledge-based mistakes in condition I and 32% of those errors in condition II were not detected at all.

Another study of error self-detection was conducted by Sellen (1990) who asked seventy-five subjects to report all errors they committed throughout the day, and how they detected them. Two questions served as an initial basis for grouping the observed results: a) “What kind of information serves as the basis for detection?”; and b) “With what is this information compared?”. Sellen categorized the reported errors as slips, mistakes, and lapses. Lapses accounted for 26.2 % of all observed errors and were detected based on reminding or retrieval of information from memory through external associations, unsatisfied goal states, internal associations, or mental review. The author determined that the detection of lapses is fundamentally a different process from the detection of slips and elected to not discuss lapses further in her study. Only 15.6% of all errors in this study required another individual to intervene. Overall, Sellen described four different mechanisms that led to detection of the reported slips and mistakes.

When the individual realized the occurrence of an error based on the perception of some aspect of the erroneous action itself, the detection was termed action-based detection. Action-based detection accounted for 11.2% of the data and occurred in the context of routinely executed habit patterns that required only minimal cognitive effort. Detection was dependent on perceiving the mismatch between the action plan and the executed action. This type of slip detection required evaluation with higher level goals and intentions. It was not always possible to immediately identify the specific error, even though there was an awareness that an error had occurred. Feedback is required for action-based detection and involves multiple forms. Detection of these errors could occur before, during, or immediately after committing the error.

Outcome-based detection occurred in 39.5% of all errors. This detection method is based on the observation of perceptual or conceptual violations of what the individual expected. Detection could also be the result of a comparison to a familiar error pattern or of the failure to achieve a goal state. A mismatch between the intention and action was

not always initially strong enough, or sufficiently monitored, to signal an error. The intention itself might also be wrong, therefore no mismatch between the action plan and executed action existed. In both cases, the action taken was not detected as being in error. Instead, it was the result or outcome of the action that was the triggering cue.

According to Sellen, an individual must be aware that the expected and actual outcomes are different in order to detect an error. This can only occur when plans and actions carry expectations about their outcomes, these outcomes are observable, the state of the world is sufficiently monitored, and the individual relates expectations to their observations.

In Sellen's study, limiting functions led to the detection of 7.6% of all errors. Limiting functions can result in error detection when constraints of the external world do not allow the initiation or continuation of a planned erroneous action.

A small number of empirical studies has looked at error detection in operational settings, including hospitals, plant and ship control rooms, and aviation. One of the earliest operational studies was conducted by Barker (1962) who investigated errors of medication administration by nurses. A combination of approaches was used including observations, self-report questionnaires, and evaluation of incident report records. The study compared the frequency of reported errors in an incident database to observed incidents.

The 93 observed errors were grouped into six medication error categories according to their phenotype. The first category was omissions (37%) described as any medication dose that was not given by the time the next does (if any) was due. The next category was wrong dosage either above (8%), or below (13%), the correct dosage by more than five percent. The third category was extra dosage given (10%), which was any does given in excess of the total number of times ordered by the physician. Unordered drug given (18%) was administration of any medication not ordered for that patient. Fifth, the wrong dosage form (4%) was any dosage form that was not included in the generally accepted interpretation of the physician's orders. The sixth category was wrong

time (10%) which was any drug given 30 minutes, or more, before or after it was ordered up to the time of the next does of the same medication.

Overall, this study concluded that undetected errors were much more prevalent than believed. Some of the medication administration problems stemmed from a lack of a built-in cross-check procedure. Errors were easily compounded through the use of a medication card to update patient records. With the use of the card system, there were no further comparisons to the original physician orders. In addition, complications displayed by patients as a result of inadequate medication were attributed to the patient illness, non-responsiveness, placebo effect, or medication reaction. By extrapolation of the error observation period to a full year, the investigators estimated that approximately 51,000 errors occurred during a year, compared to 36 actual filed reports.

An interesting finding of this study is that all 93 medication errors were detected by the researcher's confederate. The nine observed nurses in this study failed to notice their errors and consequently did not report them. In one case, a nurse committed 8 errors and was not aware of any of them. The author concluded that there is a particular tendency to miss and therefore under-report omission and timing errors.

Van Eekhout (1981) conducted a more controlled study of 36 marine engineer officers in the simulator of a supertanker engine control room. Errors and error detection were studied by means of verbal protocols, computer logs of discrete events, interviews, observations, and questionnaires. Subjects had to handle failure or fault conditions.

86 errors occurred with respect to five different sub-tasks - observation of system state, identification of fault, choice of goal, choice of procedure, and execution of procedure. The errors were classified as incomplete execution of procedures (27%), which included omission and out of sequence steps; inappropriate identification of the failure including both false acceptance and false rejection (26%); and incomplete observation of the state of the system prior to forming a hypothesis of the cause of the observed symptoms (13%). Overall, this study indicated a high frequency of omission errors as well as mistakes and further supports the notion that operators sometimes miss an error due to a partial overlap between their expectations and their observations. In

other words, they tend to complete a task given the available confirmatory evidence without searching for or attending to additional (possibly contradictory) information.

More directly related to our work are two studies that used incident databases and explored errors and error detection in the aviation domain. Wiegmann and Shappell (1997) used a database of U.S. Navy and Marine Corps aviation accidents, and Degani et al. (1991) used the Aviation Safety Reporting System (ASRS) as a source of data.

The goal of Wiegmann and Shappell's (1997) study was to explore how well three different information and human error taxonomies could be applied to the analysis of an existing database. The study used the four-stage model of information processing proposed by Wickens and Flach (1988), the model of internal human malfunction derived by O'Hare (1994), and Reason's (1990) model of unsafe acts. Wiegmann and Shappell found that they were able to classify 86.9% of the observed errors using the information processing model, 91.3% using the model of internal human malfunction, and again 91.3% using the model of unsafe acts.

Of particular interest was the distribution of errors within the different taxonomies. For the four-stage model of information-processing, errors in response execution were the most frequent (45.5%), followed by decision or response selection errors (29.5%). For O'Hare's model, procedural errors were the most frequent (39.5%), followed by diagnostic errors (21.7%). And finally, using the model of unsafe acts, Wiegmann and Shappell found that intended actions accounted for 74.5% of errors with the largest percentage of those being mistakes (57.1%).

Weigmann and Shappell also examined what error types were most often involved in major (cost of $1,000,000; total loss of aircraft; or fatality) verses minor (cost between $10,000 and $200,000; or loss of one workday) accidents. Both types of accidents were associated most frequently with response execution. Decision or response selection errors were more frequently associated with serious accidents (34.8%) than with minor accidents (24.6%). Major accidents most often involved goal (15.1%) and strategy (14.3%) errors while minor accidents were again more often associated with procedural errors (44.9%). And finally, both major and minor accidents most often involved

mistakes (57.9% and 54.9%) – an important finding for the context of our study since it suggests the need for improved support of mistake detection.

Degani et al. (1991) used the ASRS database to compare errors and error detection on traditional aircraft with those on modern automated flight decks. The study investigated who was responsible for error detection and what subsystem or information enabled error recovery. The investigators found that many sources of information were used to detect reported altitude deviations with the majority (180 out of 371) being detected by Air Traffic Controllers (ATC). Of the remaining incidents, the crews on the automated flight decks were more likely to detect an altitude deviation (approximately 100 out of 186) than those on the conventional aircraft (approximately 70 out of 185). In both cockpit types, the pilot flying was more likely to detect the deviation than the pilot not-flying (104 verses 77). They found that ATC, the altimeter, and the outside scene were the three most frequent triggers that there existed an altitude deviation.

This study did not address some important aspects of error detection. For example, it did not determine the performance level at which the crew was functioning when the error occurred. In addition, only altitude deviations were investigated.

Summary

Human error has only recently become a topic of interest in its own right. For decades, it has been studied only as a means to an end, as a way to understand normal cognitive functioning. Today, errors are being studied extensively, and numerous error classification and performance level schemes have been developed. Still, the area of error detection has received very little attention. This research is an attempt to make progress in our understanding of the relationship between errors and error detection mechanisms and of possible ways of better supporting error detection in the interest of further increasing safety in a variety of domains.

What is known to date is largely based on the work reviewed in the preceding sections. Norman has advocated the distinction between slips, lapses, and mistakes which Reason related to skill-, rule-, and knowledge-based behavior. Reason suggests that skill-based errors - slips and lapses- occur frequently and are detected by the individual

quickly and effectively. Errors related to problem solving (i.e., rule- and knowledge-based mistakes), on the other hand, tend to require intervention by an external source. Daily occurrence of problem solving behavior is less frequent than routine actions, and therefore, there are fewer opportunities for mistakes.

Sources of error detection include feedback from an action, self-monitoring, comparison of an outcome to an intention, environmental cues such as limiting functions, and the intercession of another individual. Error detection behavior has been categorized as direct error hypothesis, standard checks, and error suspicious.

Several studies have investigated the frequency of different types of errors and how these errors are detected. Allwood as well as Rizzo et al. found that subjects were able to detect one type of execution error - slips - quite frequently while mistakes were more difficult to detect. Sellen studied everyday errors in an attempt to broaden the expression of errors. She chose to discount lapses in the evaluation because the detection of lapses was considered fundamentally different from the detection of slips and mistakes. For those errors, slips and mistakes, detection based on outcome was found to be the most frequent detection mechanism. This study by Sellen represents a useful starting point for the exploration of error detection in complex domains. Barker's and Van Eekhout's research indicate a high frequency of omission errors, or lapses, and of mistakes in complex operational environments.

The aviation database studies provide a basis and direction for comparison with the work performed in this study. Wiegmann and Shappell's study successfully demonstrated the applicability of different models and categorizations of error to an existing database. They found that accidents most often involved mistakes. Degani et al. found that an external agent, ATC, was required to detect errors leading to altitude deviations. The study did not explore the relationship of the errors to performance level, nor were the errors related to general error types to allow direct comparison with other studies.

Predictions

Based on existing theories and models of human error and error detection (e.g., Reason's GEMS model) and the findings from the above discussed studies, a number of predictions can be made. These predictions concern expected error types, their relation to error detection sources and mechanisms, and the possible impact of modern technology on the nature and detection of errors such as the ASRS incidents that are being analyzed in this research, and that call for improved detection support.

Expected Frequencies of Different Errors Types in the ASRS Database. We can think of the ASRS database as a collection of reports describing episodes in which error detection was successful in the sense that an accident was prevented. At the same time, error detection occurred fairly late – late enough to lead to a potential or actual violation of regulations which is the reason these incidents were reported. This suggests that most reported cases will involve deviations from ATC-assigned or regulated limits and targets (e.g., altitude deviations). It also means that, in terms of the underlying problems, we can expect a relatively large percentage of errors in the ASRS database (relative to the likelihood of their occurrence) to be lapses/errors of omission and mistakes which are one form of error of commission. Earlier research has shown that another form of commission error – slips, tend to be detected (and corrected) rapidly and effectively by the operator committing the error before any violations and thus detection by ATC can occur (e.g., Smith, (1979)). Consequently, slips may be less likely to find their way into the ASRS database. Finally, skill-based errors will probably be the most frequently reported error type since all actions tend to have skill-based components for the implementation of any control directive (Reason, 1990). Thus, there are far more opportunities for this type of error to occur.

Most skill-based errors that appear in this database are expected to be associated with attentional failures. Since the expected outcome associated with skilled-based performance is very clearly specified, the failure to notice a discrepancy between desired and actual outcome is likely to result from a failure to attend to the corresponding feedback in the first place.

The Relation Between Error Type and Error Detection Likelihood/Mechanism. The detection of both lapses/errors of omission and mistakes is difficult for the operator committing the error (Van Eekhout, 1981; Rizzo et al., 1987) and is therefore likely to require some external intervention. In the case of mistakes, Air Traffic Control (ATC) and the other crewmember are expected to be the most frequent source of error detection because pilots, for the most part, do not form their own intentions. Instead, goals and targets are given to them by ATC. Earlier research has shown that air-ground communication of these goals and targets often breaks down (Monan, 1986; Beaty, 1995). Misunderstandings between ATC and the pilot can lead to misperception of controller intent. It is impossible for the pilot who is acting in accordance with the (assumed) controller intent to detect the mismatch between his/her actions and the controller's actual goals. ATC, however, knows about both intended and actual aircraft behavior and can therefore detect an error and point it out to the pilot. The other crewmember may also catch the error based on his/her ability to listen in on ATC communication and thus realize the other pilot's mistake.

The detection of lapses/errors of omission is expected to require salient system feedback that either captures the pilot's attention in the absence of information search or that "pops-out" when the operator performs a check on progress toward his/her goals. In addition to basic, clearly indicated flight parameters, this feedback may take the form of forcing functions, which do not allow the behavior to continue until the problem has been corrected, and/or alarms, which can capture the pilot's attention.

New Technology And Error (Detection). Errors can not be considered independent of the context in which they occur since they are often the result of a mismatch between human, system, and/or task domain. This implies that changes in any one of those elements can change the nature and/or frequency of errors. For example, in the aviation domain, the last two decades have seen the introduction of many new highly automated systems to the flight deck. Research on pilot interaction with these new

systems suggests that this technology does, in fact, create the opportunity for new kinds, of and a different likelihood, of errors (Woods et al., 1994).

One possible change in the type of errors is a shift towards errors of omission/lapses. This prediction is based on the fact that modern technology tends to operate at a high level of autonomy and authority. As a result, there is an increased likelihood that a system initiates an action without pilot input and, potentially, without pilot awareness. Consequently, the pilot may fail to notice if the machine action is inappropriate, and he/she may fail to intervene with the automation activities - an error of omission. On conventional aircraft, such errors are much less likely since systems on these airplanes are for the most part reactive in nature, i.e., they do not take an action unless and until it is explicitly commanded by the pilot/crew. Consequently, one could expect to see a much higher percentage of errors of commission.

One can also expect to see a larger percentage of mistakes on automated aircraft. Mistakes refer to both errors in the formation of an intention and to the inappropriate choice of a method for achieving a goal. The latter case may be more likely on modern aircraft since automation technology has increased the number of options available to pilots in order to achieve the same goal. For example, some automated aircraft provide the pilot with five or more different modes for changing altitude. This increased number of options affords more opportunities for choosing the wrong method or stategy and thus making a mistake.

Finally, on conventional aircraft, the pilot-flying alone was in charge of maintaining the intended or ATC-given flight path. On more automated aircraft, however, the pilot-not-flying has taken over some of the tasks involved in flight path management. For example, many airlines require the pilot-not-flying to set the altitude target for the automation. This implies that are more opportunities for the pilot-not-flying on modern aircraft to commit errors that may lead to deviations from, and violations of, ATC clearances. A related prediction associated with modern technology aircraft is that it can be more difficult for a pilot to detect an error made by the other crewmember. This prediction is based on the fact that both pilots can interact with the automation and set up

the system without the other crewmember necessarily being able to or being supported in observing the input to the automation.

In order to examine the accuracy of our predictions concerning error types, error detection, and the impact of automation technology on both factors, we analyzed ASRS incident reports to identify a) the genotype of error underlying the reported problem; b) the performance level at which the operator was functioning; and c) the cue or mechanism that led to the detection of the error. We also compared incident reports filed by pilots flying conventional aircraft versus highly automated airplanes with respect to the above issues.

Methods

ASRS Incident Data Base

The data that were analyzed in the course of this research were obtained from the Aviation Safety Reporting System (ASRS). ASRS was established as a joint cooperative program between the Federal Aviation Administration (FAA) and NASA in 1975. The ASRS mandate is to identify and report deficiencies in the National Aviation System (NAS), contribute to formulation of NAS policy, and support aviation human factors research. The ASRS data base consists of voluntarily submitted aviation incident reports. Incidents are defined as an occurrence or condition that is, or is potentially, unsafe. An incident does not involve personal injury or significant property damage (Chappell, 1994).

ASRS reports may be filed by anyone involved in, or observing, a situation in which aviation safety actually was, or could have been, compromised. The major incentive for filing a report is that none of the submitted information is used against the individual for enforcement actions. Additionally, fines and penalties are waived, subject to certain limitations, for unintentional violations of federal aviation regulations. The reporter must submit the information to the ASRS within ten days of the incident to be eligible for a waiver.

Since the inception of the program in August 1975, more than 300,000 incident reports have been submitted. While commercial aviation pilots file approximately 70% of the reports, flight attendants, air traffic controllers, mechanics, and ground personnel are also encouraged to submit reports. Each report that is submitted to ASRS is evaluated by at least two subject matter experts, pilot or air traffic controller, for safety issues. Once analyzed, the reports are de-identified before being entered into the database. This allows for confidentiality of the reporter and the organization with which they are affiliated. Once the reports have been reviewed and de-identified, they are made available to outside researchers and other interested persons upon request.

The type of information that is available in the database includes several key components related to the incident. First, the reporter is asked for some background information, including:

- Who reported the incident and what is the background of the reporter (e.g., total flying time, ratings)
- Type of aircraft involved in the incident
- Conditions in which the incident occurred (e.g., weather, airspace, location)
- Air traffic control facility involved in the incident
- Operator and mission of the flight
- Flight plan, flight phase, and control status of the flight

The reporter is then asked to provide a detailed description of the event. This outline should include information on what caused the problem in his/her opinion, what could be done to prevent its reoccurrence or correct the problem, how it was discovered, contributing factors, corrective actions, perceptions, judgments, decisions, and actions (see ASRS Reporting Form, Appendix A).

The use of incident data for studying human error involves a considerable number of benefits and serves a variety of purposes. It can further support and expand on findings from simulator studies and provide guidance for more controlled studies. For example, Orasanu and Fischer (1997) used ASRS reports in conjunction with a simulator study to investigate decision-making by aircrews. Their reason for using the incident data was to explore decision events that may not have been part of, or evolved during, the missions the crews were experiencing in the simulator. The ASRS data did, in fact, yield three additional types of decision processes that were not observed in the simulator. Chou, Madhavan, and Funk (1996) used ASRS reports to support the results of an analysis of National Transportation Safety Board (NTSB) accident reports and to provide directions for a follow-up simulator study. The focus of their work was cockpit task management (CTM) and its contribution to flight safety. The ASRS reports helped to avoid biases due to the limited set of accidents in the NTSB study, and they provided

further operational support for the importance of certain factors such as the criticality of flight into terminal areas, that were subsequently confirmed in a controlled, simulated environment.

Incident data are a rich source of information regarding reasons for, and conditions favoring, a wide range of errors and error detection mechanisms in real-world operational environments. Incident data can be used to examine hypotheses about human error that were developed in more controlled laboratory settings. The voluntary and confidential nature of the reporting system used in this study promotes operators' candid disclosure of all factors and aspects related to the incident without fear of retribution. Incident databases can be used to identify trends in the nature and severity of errors that evolve over longer periods of time. And finally, incident reports have high ecological validity – they represent reports of naturally evolving situations that occur and tend to be reported in the context of a real-world environment by highly experienced practitioners in that domain (Chappell, 1994).

Like any other form of research data, incident reports also involve some limitations. For example, there is an inherent possibility of biases regarding the type of pilot who will file a report, and the type of incident that will be reported (Wickens and McCloy, 1993). Pilots who have more to "lose" from not reporting an incident (such as commercial pilots who may lose their license and thus their source of income), are more likely to file a report to gain immunity than, for example, general aviation pilots who are flying for entertainment purposes only. Incidents that resulted in an observable deviation or violation (e.g., altitude deviations) are more likely to be reported than errors that were detected and corrected before leading to a problem. And the total number of reported incidents in the database probably underestimates the actual frequency of problems and errors since it is far more likely that an operator will not report an incident as opposed to an operator fabricating an incident that did not occur (Wickens and McCloy, 1993; Wickens, 1995).

Another potential problem is that operators reporting an incident are not necessarily trained in psychological constructs and may leave out important information. As a result, researchers sometimes have to infer what happened and what the

chronological flow of events were in the incident in order to gain valuable insight into the behavioral and contextual setting within which the operator was functioning (Harle, 1994). Also, retrospective reports are rarely completely accurate in terms of exact details and the chronological flow of events (Loftus, 1979).

Selection Criteria For Limiting The Database

The entire ASRS database contains over 307,000 incidents. However, complete detailed reports are available for only 58,021 incidents that occurred between January 1988 and May 1996. From these reports, the following selection was made. The mission profile was limited to commercial passenger flights since safety improvements in this area can be considered particularly important. Changes can save lives and improve the public's perception of this mode of transportation. Given that one of our interests was a comparison between conventional and highly automated aircraft, eleven aircraft types were selected which included six advanced technology aircraft (Airbus A-320, McDonnell-Douglas MD-11, Boeing B-737-300, B-757, B-767, and the B-747-400) and five conventional aircraft that do not involve high levels of automation (the McDonnell-Douglas MD-80, the DC-9, the DC-10, the Boeing B-737-200, and the B-747-200). These aircraft were chosen since they represent pairs that differ only in terms of their level of automation while aircraft size and routing are comparable. The pairs are B-737-200 and B-737-300, DC-10 and MD-11, B-747-200 and B-747-400, DC-9/MD-80/A-320, and the B-757 and B-767.

In our data analysis, we started with the most recent incidents (May 1996) and worked backwards to the last month in the database (January 1988). Overall, 1091 reports fit our profile and were reviewed. Of these, only 245 reports (22%) could be included in the final data analysis. The number of incidents that we were able to include for each aircraft is shown in table 3.

Aircraft	Number	Aircraft	Number
B-737-200	37	B-737-300	37
DC-10	11	MD-11	5
B-747-200	6	B-747-400	3
DC-9 / MD-80	36 / 34	A-320	24
B-757	26	B-767	26

Table 3: Aircraft Type and Number of Incidents Included in Data Analysis

The remaining reports were excluded for the following reasons: a) Reports that did not involve a specific operator error were eliminated. These included general problems or warnings such as a poor lighting system or difficulty understanding a controller at a particular airport; b) Reports of incidents that were beyond the reporter's control (e.g., bird strike, mechanical problem) were also eliminated; c) Incidents resulting from intentional violations were not of interest since they do not require or involve the detection of an error; d) Reports that were filed regarding another aircraft or filed by an individual who was not a member of the cockpit flight crew were eliminated since it was not possible to reliably determine whether those reports accurately reflected what had happened in the incident; e) Finally, a large number of reports had to be excluded since they did not explicitly state the factors or mechanisms that led to the detection of the error.

Data Analysis

The remaining ASRS reports were analyzed using a form that we designed specifically to gather information concerning the questions raised in the introduction to this document (see Data Analysis Form, Appendix B). This form captured the following aspects of the incident:

- Who committed the error

- Who detected the error
- A brief description of the incident
- Error phenotype
- Classification of the error
    - Error of omission or commission
    - Slip, mistake, or lapse
- Type of task performance level at which the error occurred
    - Skill-, rule-, or knowledge-based
- Contributing factors, (e.g., inattention; time pressure; distraction)
- Error detection source and mechanism (e.g., self/other operator/ATC; routine/suspicious check; limiting function, etc.)

Several passes through the database were made. Each incident was first analyzed to determine whether it involved an omission or commission error. The same report was then reviewed to determine if the erroneous action was a slip, lapse, or mistake. Finally, the error involved in the incident was analyzed to determine the performance level at which it occurred - - skill-, rule-, or knowledge-based behavior. These categorizations were performed independently of each other. We chose to analyze the data using the various classification schemes because some of the categories in one scheme (e.g., commission error; mistake) include more than one kind of error from a different scheme (e.g., slips and mistakes, and rule- and knowledge-based errors respectively). Use of only one scheme could have hidden interesting differences and effects.

The following are abbreviated examples of slips, lapses, mistakes, and of skill-, rule-, and knowledge-based errors from our database:

Slip: During the approach, the aircraft was cleared to descend to and maintain 8,000 feet. Even though the Captain understood the clearance and meant to set 8,000 feet in the altitude alert window, he inadvertently entered 3,000 feet - - a slip.

Lapse: The aircraft was cleared to climb to 10,000 feet. During the climb the Captain, who was the pilot flying, became distracted by the actions of the First Officer and failed to level-off at the assigned altitude although he had intended to - - a lapse.

Mistake: This was the second landing for the Captain after initial orientation. He was trying to avoid flaring too early and was therefore waiting for the 30 and 20 foot callouts by the automation. He did not realize that these callouts do not occur if the aircraft descends through the altitudes too quickly. The result was a very hard landing. This is an example of a mistake where the pilot's intention was inappropriate for the given situation.

Skill-based Error: Both the slip and lapse examples above are errors involving skill-based behavior. In each case, the operator is performing a routine activity in a familiar situation, but the execution has broken down. Their intentions are appropriate; however, they error in the execution of those intentions.

Rule-based Error: The mistake above is also an example of a rule-based error. In this case, it involves the misapplication of a good rule. The rule is to wait for the 20 foot callout before initiating the flare, but in this case the rule is applied in the wrong context.

Knowledge-based Error: The Captain determined that the approach was unstable, and he initiated a go-around. The flight directors were off, and the aircraft was at an altitude below 100 feet. In those conditions, the automation disconnects the autothrust when a go-around is initiated. The Captain, who did not know about this aspect of system behavior, did not select climb thrust and re-engage the autothrust system. As a result, the aircraft was still at full thrust as he began to level-off. The aircraft oversped, and the crew initiated a late turn back due to excessive airspeed. In this case a novel situation is encountered, and a lack of knowledge and understanding of the system leads to the error.

Once the incident had been analyzed in terms of its underlying error types and performance level, the error detection source and mechanism/cue was determined. We first identified who detected the error: the operator committing the error, the other crewmember, ATC, or ground personnel. Next, we determined the mechanism or cue that led to error detection. The mechanisms/cues included:

Routine checks: At several points during the flight, the Captain performed cross-checks of the settings in the Flight Management System. During one of those regular checks, he realized they would not make a crossing restriction.

Suspicious checks: The First Officer began to doubt the validity of the position report that the crew had given to ATC. He began to investigate the situation and found a problem with the clock setting.

Alarms: The pilot was hand-flying the airplane during the final descent when he diverted his attention to check the taxiway he was going to use. The altitude alert sounded as the aircraft descended below the set altitude.

Limiting functions: The First Officer tried to enter a new restriction in the Flight Management System. The system would not accept the information. After another attempt the First Officer discovered he was trying to use an incorrect mode.

Outcome of an action unrelated to aircraft performance/behavior: The pilot tried to preselect a radio frequency on the second radio channel. He inadvertently changed the frequency of the active channel, and the crew noticed the error when the active channel suddenly went quiet.

Aircraft performance/behavior: The Captain continued to hold the airspeed recommended by the Flight Management System even when he disconnected the

autopilot to initiate a turn for the approach. Since the airspeed was too slow for that maneuver, the aircraft began to buffet as it approached a stall

Each incident report was independently analyzed by two researchers. Any discrepancies in their analyses were resolved through discussion until an agreement was reached. Non-parametric statistical analyses were performed for the entire sample of incidents and for the comparison between conventional and highly automated aircraft.

Results

Frequencies of Different Error Phenotypes and Genotypes In ASRS Database

The phenotype of reported incidents. All 245 incidents were divided into six different categories based on their phenotype or observable outcome. There were 88 (35.9%) altitude deviations, 79 (32.2%) course or heading deviations, 25 (10.2%) taxi errors or runway incursions, 10 (4.1%) airspeed deviations, 11 (4.5%) failures to obtain a clearance prior to take-off or landing, and 32 (13.1%) other errors such as improper fuel load, improper use of a system, or not retracting/extending equipment when required (see Figure 3).

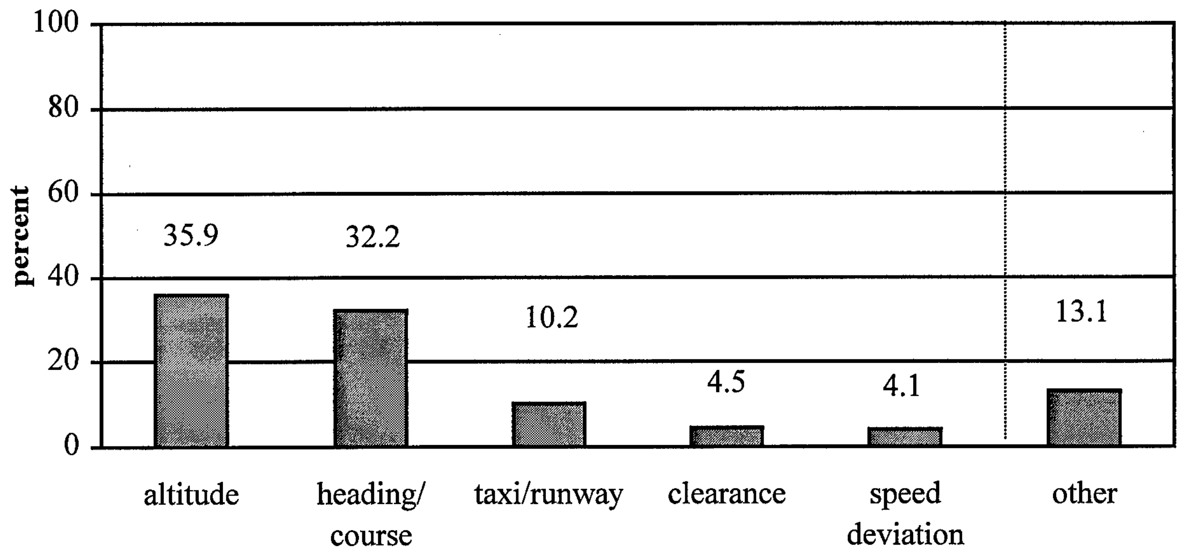

Figure 3: The Phenotype of the Reported Incidents

The above data allow for a comparison with existing accident and incident statistics and with findings from earlier research that focused on the surface appearance of errors. However, they do not provide insight into the mechanisms underlying the observed problems. To illustrate the importance of going beyond the phenotype, we selected the two most frequently reported problems, altitude and heading or course deviation, and are showing their underlying error types in table 4. These data illustrate that it is inappropriate to analyze incidents in terms of their surface appearance only. The

development of effective countermeasures to error and of error detection support requires knowledge about the underlying cognitive mechanisms and error forms.

	Slips	Lapses	Mistakes
Altitude Deviation, n = 88	17 (19.3%)	42 (47.7%)	29 (32.9%)
Heading/Course Deviation, n = 79	23 (29.1%)	26 (32.9%)	30 (37.9%)

Table 4: Different Error Types Underlying Reported Incidents

<u>The genotype of reported incidents.</u> Our first expectation regarding the nature of errors reported in the database was a high frequency of omission errors and, correspondingly, a high frequency of lapses. We also expected that, while most errors would occur during skill-based behavior, a relatively high number of mistakes (relative to opportunity for error) would be observed.

For our overall sample of ASRS reports, we found that lapses were indeed the most frequent error type, followed by mistakes. There were 49 (20.1%) slips, 101 (41.4%) lapses, and 94 (38.5%) mistakes (see Figure 4). The frequency of error types differed significantly, $\chi 2$ (2, N=244) = 19.58, p< .001.

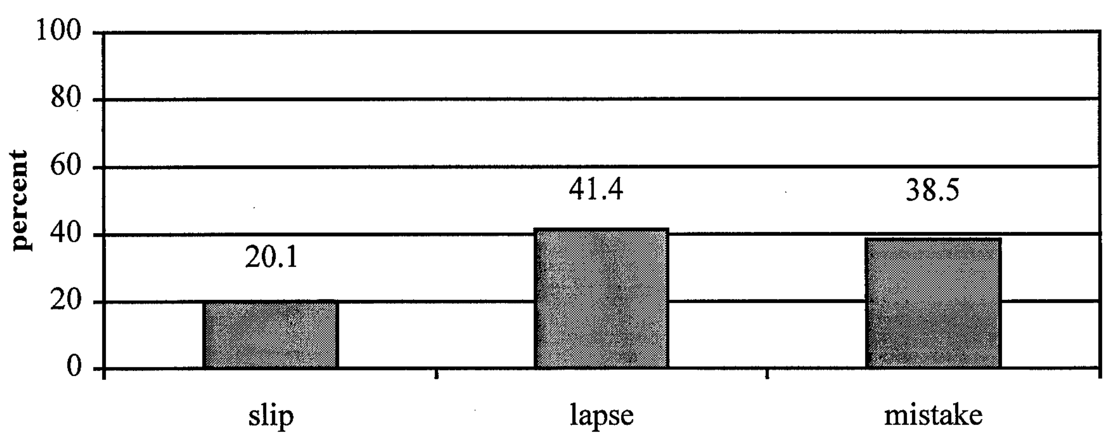

Figure 4: Frequency of Slips, Lapses, and Mistakes

As expected, the majority of errors occurred at the skill-based performance level. There were 185 (75.8%) skill-based errors, 35 (14.3%) rule-based errors, and 24 (9.8%) knowledge-based errors (see Figure 5). The difference in error type distribution was again significant, $\chi 2$ (2, N=244) = 198.94, p< .001.

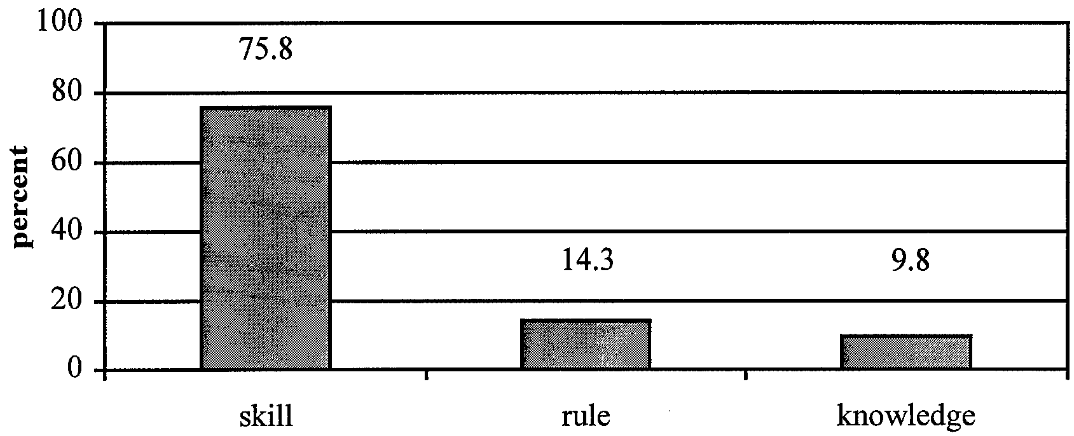

Figure 5: Frequency of Skill, Rule, and Knowledge-Based Performance Errors

Finally, we found a marginally significant difference in the frequency of omission versus commission errors across all aircraft, $\chi 2$ (1, N=245) = 3.92, p< .05. There were 107 (43.7%) omission errors and 138 (56.3%) commission errors (see Figure 6). Note that commission errors include both slips and mistakes.

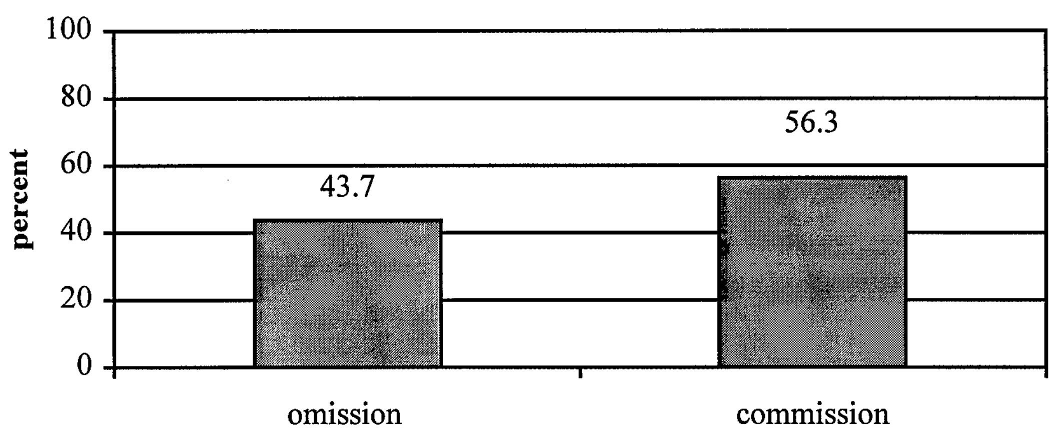

Figure 6: Frequency of Omission and Commission Errors

<u>Distractions and competing demands as major contributors to skill-based errors.</u> Previous research has suggested that one major contributor to errors at the skill-based

level is some form of attentional failure, often due to distractions or capture of some other task (Reason, 1990). Our data confirm this hypothesis - attentional problems did, in fact, contribute to 167 of the reported errors. In 86.9% of the omission errors, inattention was explicitly discussed as a contributing factor, while fewer commission errors (53.6%) were associated with attentional problems (see Figure 7). When we break down further the omission and commission errors, we find that 89.1% of all lapses and 79.6% of all slips were related to attentional failures (see Figure 7), whereas only 39.4% of the mistakes involved inattention to the task at hand.

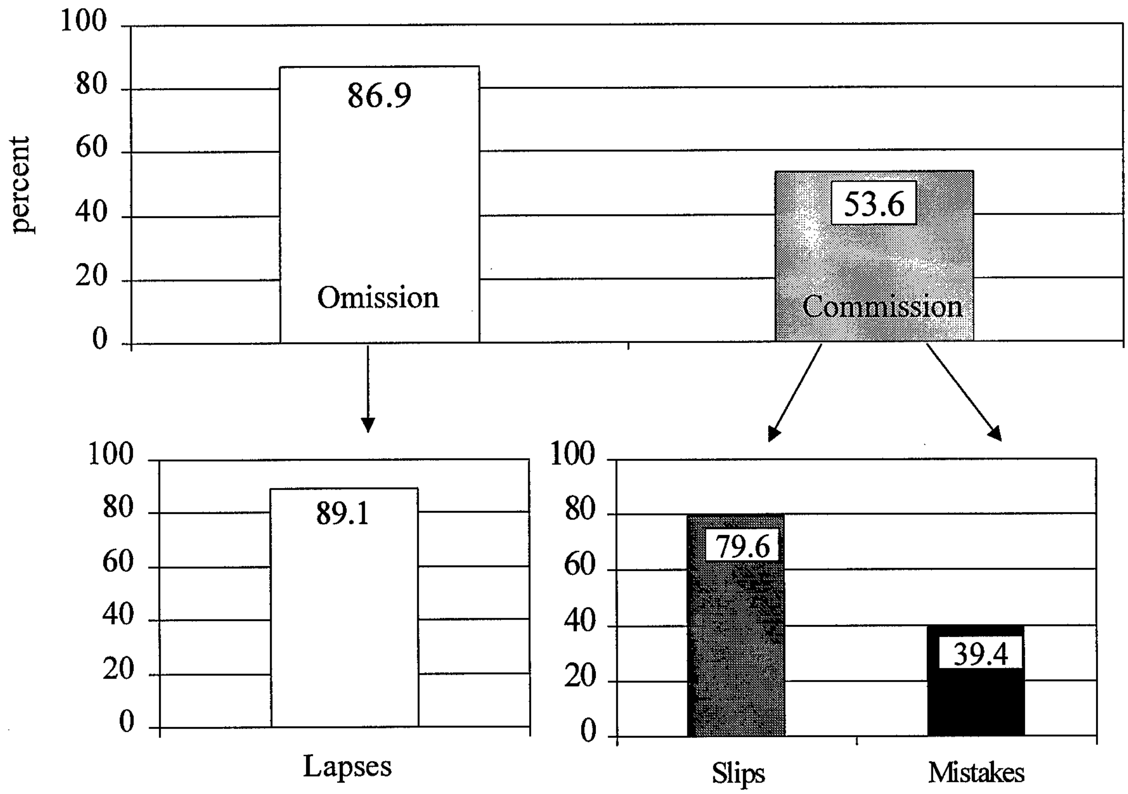

Figure 7: Percentage of Errors Related to Inattention

For the 129 skill-based errors that involved inattention as a contributing factor, the three most common sources of distraction were:

- difficulty handling equipment (malfunction, unfamiliarity) – 22 cases
- interruption (e.g. by flight attendant, jumpseat rider) – 20 cases

- competing demands and time pressure – 20 cases

Table 5 compares the slips and lapses in terms of how often they involved the above factors.

	Slips	Lapses
Difficulty handling equipment	**9**	13
Interruption	0	**20**
Time pressure	7	13

Table 5: Most Frequent Underlying Reasons For Inattention Related to Slips and Lapses

Note that only lapses were caused by interruption, whereas difficulties with handling equipment and time pressure played a role in both types of skill-based error.

<u>The Relationship Between Error Type and The Likelihood/Source of Error Detection</u>

We first analyzed the 245 incidents in terms of who detected the error. We found that 54 (24%) errors were detected by the operator him/herself, 118 (52.7%) by Air Traffic Control (ATC), 42 (18.8%) by the other crewmember, and 10 (4.5%) by ground personnel such as maintenance or dispatch (see Figure 8). There was a significant difference in the frequency of the error detection source, $\chi 2$ (3, N=224) = 110.00, p< .001.

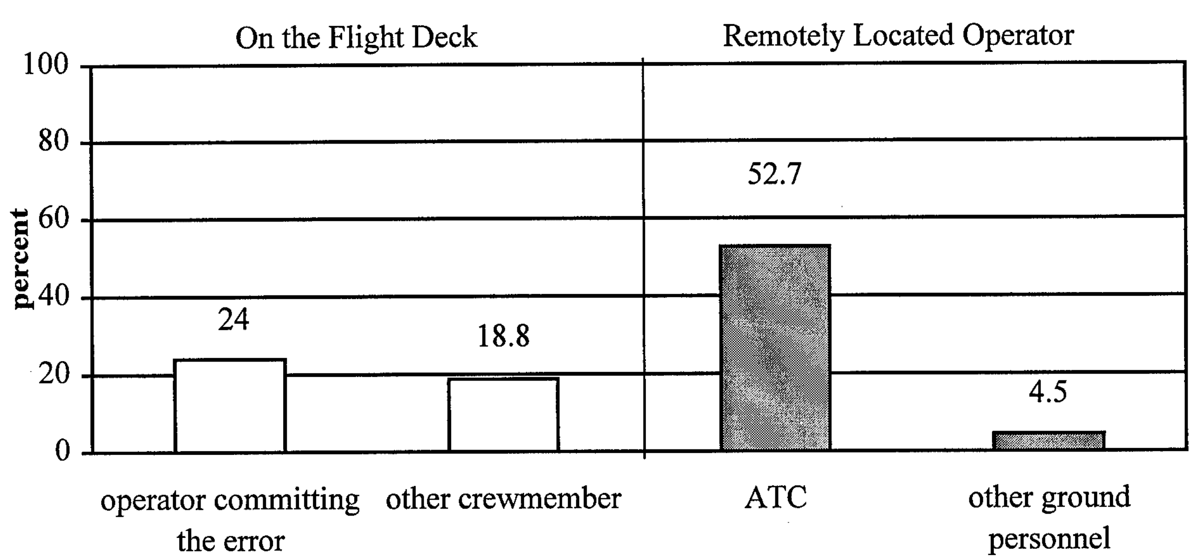

Figure 8: The Source of Error Detection - Who Detected the Error?

For errors of omission and commission, we found that ATC detected the majority of both kinds of errors (see Table 6). There was no significant difference between the detection sources for omission and commission errors.

	Omission	Commission
Operator Who Committed The Error	22 (24.4%)	23 (17.8%)
Other Crewmember	16 (17.8%)	29 (22.5%)
ATC	**47 (52.2%)**	**73 (57.4%)**
Other Ground Personnel	5 (5.6%)	3 (2.3%)

Table 6: Detection of Omission and Commission Errors

Earlier research (e.g., Reason, 1990) suggests that skill-based errors are detected rapidly and effectively by the operator committing the error. This may be too broad a prediction however. Skill-based errors include both slips and lapses, and lapses were found to be difficult to detect in earlier studies (Van Eekhout, 1981; Rizzo et al., 1987). To examine this, we identified who detected slips, lapses, and mistakes verse skill-, rule-, and knowledge-based errors.

ATC detected the majority of all slips, lapses, and mistakes in the database (see Table 7). There was again no significant difference in the source of error detection between slips, lapses, and mistakes.

	Slip	Lapse	Mistake
Operator Committed The Error	6 (14.0%)	22 (25.9%)	17 (18.9%)
Other Crewmember	11 (25.6%)	14 (16.5%)	20 (22.2%)
ATC	**26 (60.5%)**	**44 (51.8%)**	**50 (55.6%)**
Other Ground Personnel	0 (0%)	5 (5.8%)	3 (3.3%)

Table 7: Detection of Slips, Lapses, and Mistakes

With respect to the performance level, ATC detected almost 60% of the skill- and rule-based errors each, while the operator committing the error detected 50% of the knowledge-based errors (see Table 8). The difference in proportion between the source of detection for skill-, rule-, and knowledge-based errors was significant, $\chi 2$ (6, N=216) = 18.02, p< .006.

	Skill-Based	Rule-Based	Knowledge-Based
Operator Committed The Error	31 (18.9%)	3 (10.3%)	**12 (50.0%)**
Other Crewmember	31 (18.9%)	9 (31.0%)	5 (20.8%)
ATC	**96 (58.5%)**	**17 (58.6%)**	6 (25.0%)
Other Ground Personnel	6 (3.7%)	0 (0%)	1 (4.2%)

Table 8: Detection of Skill-, Rule-, and Knowledge-Based Errors

Next, we examined what cue or information supported error detection. We were not able to determine the cues used by ATC or ground personnel from the information available in the ASRS reports. However, for the 120 incidents where the operator committing the error or the other crewmember detected the error, we could identify the detection cue. Outcome of an action unrelated to aircraft performance/behavior was the basis for detection in 27 incidents (22.5%), routine checks in 37 cases (30.8%), suspicious checks in 18 incidents (15.0%), aircraft performance/behavior in 17 events (14.2%), some limiting function in 5 cases (4.2%), and alarms were involved in 16 incidents (13.3%) (see Figure 9). The difference in the frequency of identified detection cues was significant, $\chi 2$ (5, N=120) = 29.60, p< .001.

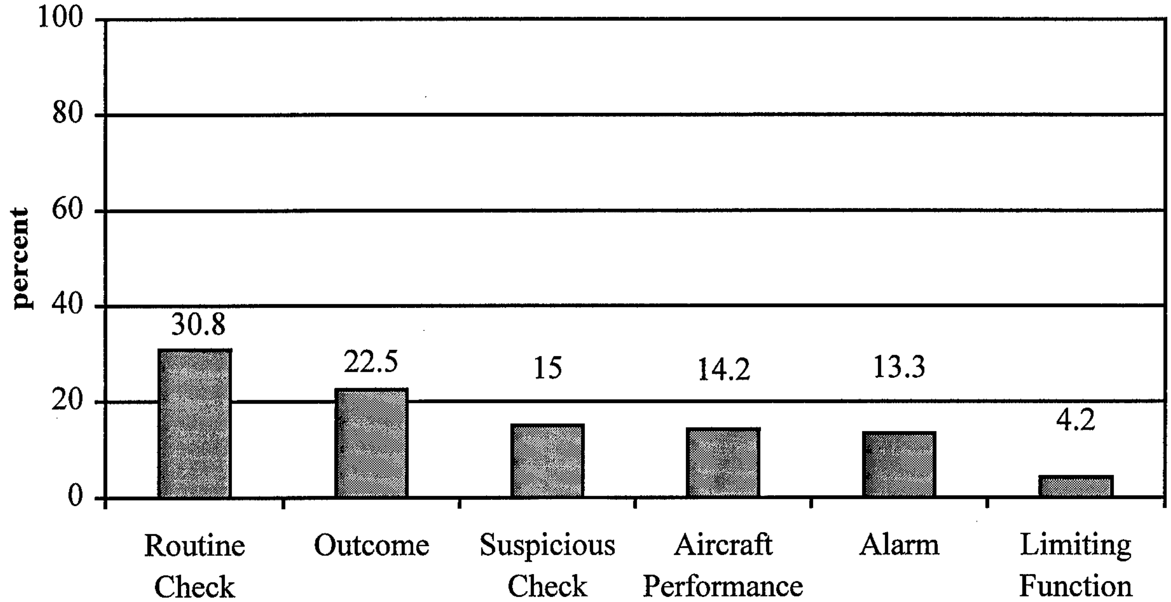

Figure 9: Cues/Mechanisms Involved in Error Detection

A more detailed analysis was performed to determine whether certain cues/mechanisms are particularly effective in the detection of different kinds of errors (see Table 9). No significant difference was found between the frequency distributions of detection sources for errors of omission and errors of commission. However, we found that not all sources of detection were equally prevalent within the group of errors of omission, ($\chi 2$ (5, N=38) = 33.68, p< .001), and the group of errors of commission ($\chi 2$ (5, N=52) = 26.46, p< .001). Routine checks (50%) were found to be the most frequent detection mechanism for omission errors, while the two most frequent sources of detection for errors of commission appear to be the outcome of an action unrelated to aircraft performance/behavior (32.7%) and routine check (28.8%).

	Omission	Commission
Outcome of action (not related to aircraft perform.)	6 (15.8%)	**17 (32.7%)**
Routine check	**19 (50.0%)**	**15 (28.8%)**
Suspicious check	6 (15.8%)	10 (19.2%)
Aircraft performance/display	4 (10.5%)	8 (15.4%)
Limiting function	2 (5.3%)	1 (1.9%)
Alarm	1 (2.6%)	1 (1.9%)

Table 9: Detection Mechanism/Cue for Omission and Commission Errors

No significant difference was found between the frequency of distributions of detection sources for slips, lapses, and mistakes. Among the detection cues for slips, we did not find a significant difference, however, the detection cues among the lapses ($\chi 2$ (5, N=36) = 27.66, p< .001), and mistakes ($\chi 2$ (5, N=37) = 22.51, p< .001) were significantly different. The most frequent detection cue for lapses was a routine check (47.2%), while the outcome of an action unrelated to aircraft performance/behavior (35.1%) and routine check (32.4%) were the most frequent detection cues for mistakes (see Table 10).

	Slip	Lapse	Mistake
Outcome of action (not related to aircraft perform.)	4 (23.5%)	6 (16.7%)	**13 (35.1%)**
Routine check	**5 (29.4%)**	**17 (47.2%)**	**12 (32.4%)**
Suspicious check	4 (23.5%)	6 (16.7%)	6 (16.2%)
Aircraft performance/display	4 (23.5%)	4 (11.1%)	4 (10.8%)
Limiting function	0 (0%)	2 (5.6%)	1 (2.7%)
Alarm	0 (0%)	1 (2.8%)	1 (2.7%)

Table 10: Detection Mechanism/Cue for Slips, Lapses, and Mistakes

Finally, we evaluated the detection mechanisms for skill-, rule-, and knowledge-based performance. No significant difference was found between the frequency

distribution of detection sources for skill-, rule-, and knowledge-based errors. However, there was a significant difference among the error detection cues/mechanisms for skill-based errors, ($\chi 2$ (5, N=60) = 41.00, p< .001). Routine checks appeared to be the most frequent detection mechanism (43.3%) in cases of skill-based errors, while the outcome of an action not related to aircraft performance/behavior (30.8%) and a routine check (30.8%) were equally frequent for rule-based errors. Knowledge-based errors were detected most frequently by the outcome of an action not related to aircraft performance/behavior (41.2%) (see Table 11).

	Skill-Based	Rule-Based	Knowledge-Based
Outcome of action (not related to aircraft perform.)	12 (20.0%)	**4 (30.8%)**	**7 (41.2%)**
Routine check	**26 (43.3%)**	**4 (30.8%)**	4 (23.5%)
Suspicious check	11 (18.3%)	3 (23.1%)	2 (11.8%)
Aircraft performance/display	8 (13.3%)	2 (15.4%)	2 (11.8%)
Limiting function	2 (3.3%)	0 (0%)	1 (5.9%)
Alarm	1 (1.7%)	0 (0%)	1 (5.9%)

Table 11: Detection Mechanism/Cue for Skill-, Rule-, and Knowledge-Based Errors

The Impact of Modern Automation Technology on Error Forms and Error Detection.

Error forms. The following analysis compares those reports in our sample that were filed by pilots on conventional (n=124) versus on automated aircraft (n=121). We first examined our hypothesis that omission errors are more likely on automated aircraft than on conventional aircraft. There were 42 (46.7%) omission errors on the conventional aircraft and 48 (53.3%) on the automated aircraft. On the automated aircraft, there were 65 (41.9%) omission errors and 90 (58.1%) commission errors. No significant difference between commission and omission errors was found (see Figure 10).

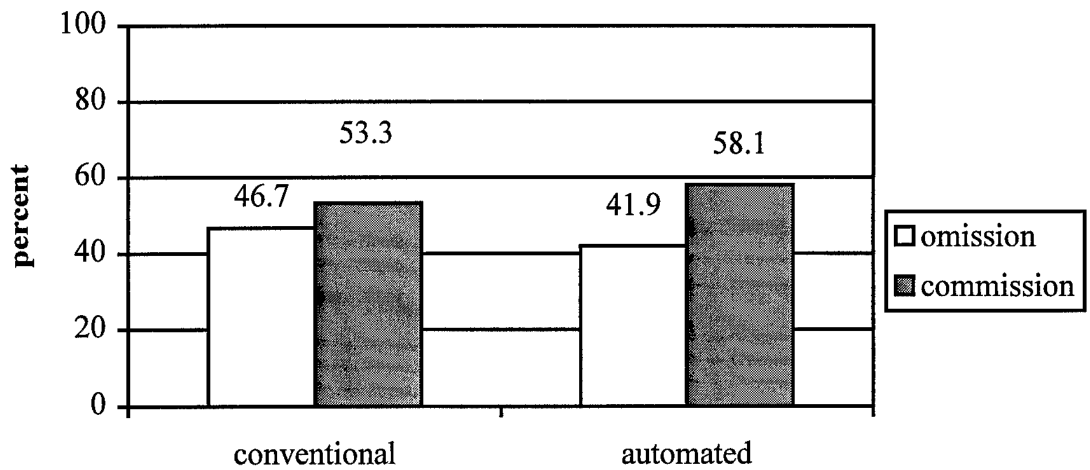

Figure 10: Errors of Omission and Commission on Automated Versus Conventional Aircraft

Similarly, no significant differences were found for the distribution of slips, lapses, and mistakes nor for errors at different performance levels (see Figures 11 and 12). There were 17 (19.1%) slips, 40 (44.9%) lapses, and 32 (36%) mistakes on the conventional aircraft. On the automated aircraft, there were 32 (20.6%) slips, 61 (39.4%) lapses, and 62 (40%) mistakes.

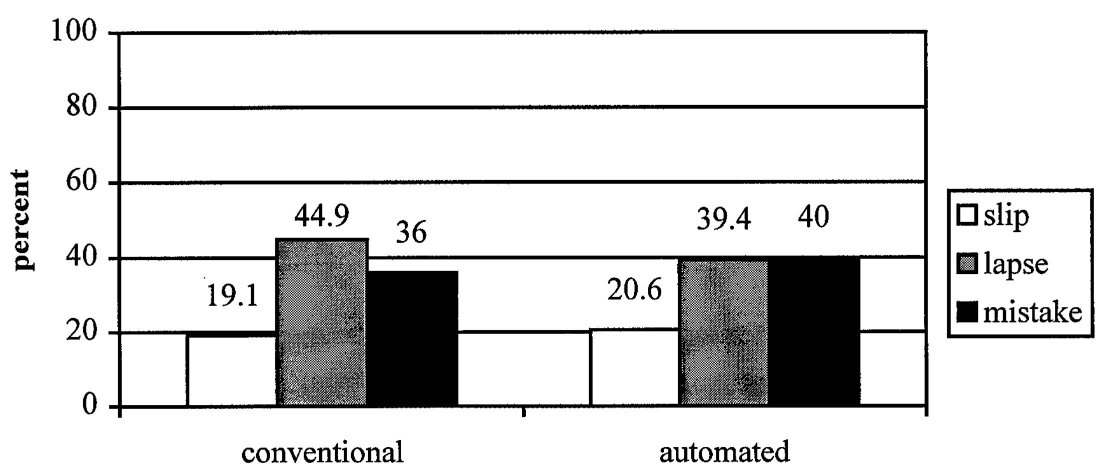

Figure 11: Slips, Lapses, and Mistakes on Automated Versus Conventional Aircraft

There were 67 (75.3%) skill-based errors, 14 (15.7%) rule-based errors, and 8 (9%) knowledge-based errors on the conventional aircraft. Pilots reported 118 (76.1%)

skill-based errors, 21 (13.5%) rule-based errors, and 16 (10.3%) knowledge-based errors on automated aircraft.

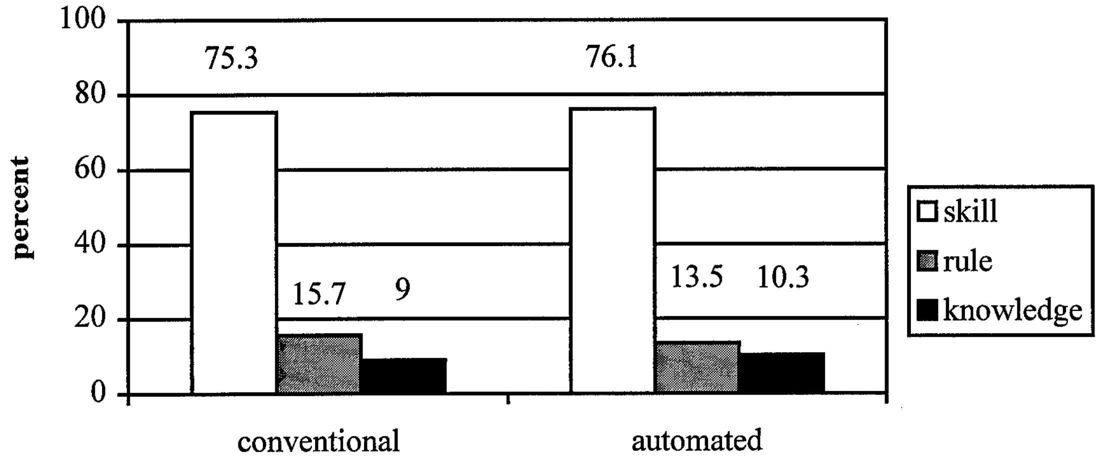

Figure 12: Skill, Rule, and Knowledge-Based Errors on Automated Versus Conventional Aircraft

It is possible that the effects of automation become visible only for errors related to flight path control (e.g., altitude, heading) since this is the primary purpose /domain of systems such as the Flight Management System. We therefore compared automated versus conventional aircraft with respect to those tasks only. But again, no significant difference was found between the frequencies of different error types.

Who is committing errors on different flight decks. Our next prediction was that the pilot not-flying on automated aircraft has more opportunities to commit errors related to flight path control than the pilot not-flying on conventional aircraft. Overall, the pilot flying was found to commit the majority of errors on both the conventional and the automated aircraft (see Figure 13). However, as anticipated, the pilots not-flying (n=33) on automated aircraft commit more errors than those on conventional aircraft (n=10), $\chi 2$ (1, N=43) = 12.30, p< .001.

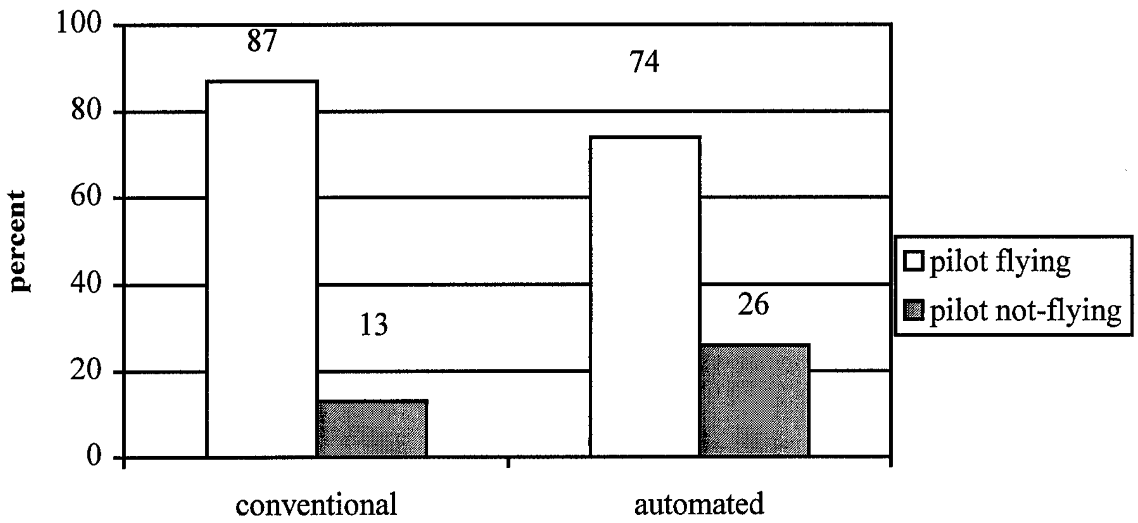

Figure 13: Error Committed, by Flight Crew Position

In particular, we found that the pilot not-flying on the automated aircraft committed considerably more commission errors ($\chi 2$ (1, N=29) = 5.83, p< .016), slips ($\chi 2$ (1, N=14) = 7.14, p< .008), lapses ($\chi 2$ (1, N=11) = 4.45, p< .035), and skill-based errors ($\chi 2$ (1, N=29) = 5.83, p< .016) than the pilot not-flying on conventional aircraft. This finding appears to be reversed for the pilot flying (see Tables 12, 13, and 14). However, in this case, the differences were not significant.

	Pilot Not-Flying		Pilot Flying	
	Conventional	Automated	Conventional	Automated
Omission	6	7	33	31
Commission	8	21	45	33

Table 12: Frequencies of Omission and Commission Errors By Crew Position

	Pilot Not-Flying		Pilot Flying	
	Conventional	Automated	Conventional	Automated
Slips	2	12	19	7
Lapses	2	9	32	27
Mistakes	4	9	28	25

Table 13: Frequencies of Slips, Lapses and Mistakes Committed By Crew Position

	Pilot Not-Flying		Pilot Flying	
	Conventional	Automated	Conventional	Automated
Skill	8	21	66	50
Rule	0	4	7	10
Knowledge	1	3	2	4

Table 14: Frequencies of Skill-, Rule- and Knowledge-Based Errors Committed By Crew Position

Detection of errors by the other crewmember on highly automated aircraft. Due to problems with observing the actions of other crewmembers on highly automated aircraft, we expected that the other crewmember (the crewmember who did not commit the error) would be less likely to detect an error. Our data suggest, however, that the opposite is the case. The other crewmember detected a greater percent of errors on the automated aircraft than on the conventional airplane, $\chi 2$ (1, N=33) = 10.93, $p < .001$(see Table 15).

	Conventional (n=124)	Automated (n=121)
Errors detected by the other crewmember	7 (5.6%)	26 (21.5%)

Table 15: Frequencies of Errors Detected By the Other Crewmember

Discussion

There is growing concern in the aviation industry that the anticipated growth in air traffic will lead to an increased number of accidents. Since human error is cited as a contributing factor in the majority of aviation accidents, one promising avenue towards lowering the accident rate is to invest in a better understanding of the nature of, underlying reasons for, and potential countermeasures to erroneous actions and assessments. While it is not possible to completely eliminate errors, operators can be supported in detecting and recovering from them in time to avoid catastrophic consequences. One source of information about human error and error detection is incident reports which describe precursor events that did not result in accidents since they occurred in an error-tolerant environment, or were detected in time to prevent severe consequences. Investigators have, for years, argued the importance of incident investigation as a method of exploring, and possibly preventing, accidents (Fitts and Jones, 1947; Heinrich, 1980; Diehl, 1991). In this study, we analyzed incidents reported to the Aviation Safety Reporting System in terms of their underlying error types (omission/commission errors and slips/lapses/mistakes) and performance level (skill-, rule-, or knowledge-based). We then examined how these errors were detected - both in terms of the source of detection and the cue or mechanism involved. Finally, the potential impact of modern automation technology on the nature of errors and error detection was explored.

The Frequency of Different Error Phenotypes and Genotypes

We found that, when analyzed in terms of their phenotype or surface appearance, altitude and heading/course deviations were the most frequently reported problems (see Figure 3). This confirms findings from earlier research. For example, O'Hare (1990) reported a large number of directional (heading/course) deviations, especially for the take-off and descent phases of flight. Monan (1986) found that altitude and heading deviations were the most frequent outcome in his study of miscommunication and misunderstandings between pilots and air traffic control. And Degani et al. (1991)

focused on altitude deviations in their study because of the high frequency of these incidents in the ASRS database. The large number of altitude-related difficulties may reflect the absence of vertical situation displays on current flight decks.

Analyzing incidents and accidents in terms of their surface appearance alone can be inadequate, however. As illustrated by our data (see Table 4), seemingly homogenous groups of incidents may involve very different underlying errors. For example, 19.3% of the altitude deviations included in our sample turned out to be related to a slip while 46.6% of the altitude deviations involved a lapse, and 32.9 % were the result of a mistake. Identifying these underlying errors is important since they call for different countermeasures and involve different detection mechanisms and probabilities.

For the most part, our hypotheses regarding the frequencies of error types in the database were confirmed. Most incidents involved lapses and mistakes (see Figure 4) which are quite difficult to detect and therefore likely to remain unnoticed long enough to result in some problem or violation. Slips, on the other hand, were expected and found to be involved in only 20.1% of the incidents. They tend to be detected fairly rapidly (Reason, 1990) and are therefore unlikely to make their way into the ASRS database. This assumption is supported by Smith's (1979) and Barker's (1962) findings that far more errors occur in a variety of domains, than are ever reported since the errors are corrected immediately. Our data confirm earlier findings by Wiegmann and Shappell (1997), and Woods (1987), and suggest a considerable need for better support of detection of lapses and mistakes. Currently, many of these errors are caught by the final layer of defense in the overall system – a situation that is not desirable.

As expected, most reported errors occurred when the pilot was operating at the skill-based performance level (see Figure 5). This can be explained by the fact that "virtually all adult actions … have very substantial skill-based components." (Reason, 1990). In other words, there are far more opportunities for skill-based errors and thus the absolute number of those errors can be expected to be high even though the ratio of error to opportunity may be lower than that for rule- and knowledge-based errors.

Note that the classification of errors as skill-, rule-, and knowledge-based is problematic. Like the category of errors of commission, skill-based errors include two

very different error types - slips and lapses. Our expectations for these two error types were different. We anticipated relatively many lapses but few slips. When looking at errors at the performance level, however, this difference is not visible as both error types fall under the label "skill-based". This affects the interpretability of earlier findings. For example, averaging over a number of studies by Allwood (1984), Bagnara et al. (1987), and Rizzo et al. (1987), Reason (1990) points out that 86.1% of all skill-based errors in those studies were detected by the operator. This does not provide any insight into whether slips or lapses are equally likely to be detected. Allwood's study in particular demonstrates that skill-based errors can not all be claimed as readily detectable (Reason, 1990).

We decided to use the skill-, rule-, knowledge classification in our data analysis despite the above shortcoming because it also involves a potential benefit. It allows us to distinguish between different types of mistakes in our database. Mistakes can take the form of rule- or knowledge-based errors – two types of error that occur at different levels of performance and may therefore differ in terms of their likelihood and ease of detection. With one exception (see Table 8), however, no significant differences were found between the two error types. The one exception involves a significant difference between the frequency distributions of the sources of error detection for rule- and knowledge-based errors. Air traffic control was the most frequent source of error detection for rule-based errors whereas the pilot committing the error most often detected knowledge-based errors. This way may be related to Reason's (1990) claim that rule-based errors are of the "strong-but-wrong" kind. In other words, the operator making a rule-based error tends to be quite convinced of the appropriateness of his/her actions since the actions are based on pre-existing, well-established rules. Therefore, he/she fails to check for contradictory evidence. In contrast, knowledge-based errors occur during on-line problem solving based on a trial-and-error approach that is more likely associated with some degree of uncertainty. This uncertainty may cause the operator to more actively search for information on whether or not their actions were successful in achieving the desired goal or solution. Consequently, knowledge-based errors may require external intervention less often than rule-based errors.

The results presented in table 11 suggest another trend related to the detection mechanisms/cues associated with these two error types. Knowledge-based errors appear to be detected most often by the outcome of an action unrelated to the aircraft performance/behavior whereas rule-based errors are detected equally often based on action outcome unrelated to aircraft performance/behavior and routine check.

Breakdowns in skill-based performance are assumed to result most often from attentional failures due to inattention, i.e., failing to make a necessary check, or misallocation of attention, i.e., making an attentional check at an inappropriate point in a routine sequence (Reason, 1990). This assumption was confirmed by our data (see Figure 7) which show that a considerable number of slips and lapses (89.1% and 79.6 %, respectively) - the two error forms at the skill-based level - involved attentional problems. These were related to difficulties with handling unfamiliar or malfunctioning equipment or to competing demands in high-tempo operations. Lapses also involved interruptions of a task by someone on the flight deck. These findings (see Table 5) suggest possible areas for further investigation and possible ways of reducing the number of skill-based errors. For example, distractions and interruptions may be reduced by enforcing stricter rules for sterile cockpit operations. And more effective use of cockpit resource management may help minimize attentional problems due to excessive competing demands on one operator (Chou, et al., 1996; Rouse and Morris, 1987). Distractions and inattention have been identified before as major contributing factors to errors in earlier studies of air traffic control (Maurino, 1995), daily activities (Sellen, 1990), and civil aviation (Farmer, 1994).

In summary, while most accident analyses to date focus on the phenotype or surface appearance of error, we have shown that this approach is of limited use when trying to understand and address human error. Instead, the analysis of the genotype of error is critical to identify common underlying problems and develop corresponding countermeasures. In our study, lapses and mistakes were found to be the most frequent type of error. This is in line with the assumption that these errors are the most difficult to detect and require better support of operators. A comparison of the different error classification schemes used in our analysis suggests that the most appropriate approach

may be to use the distinction between slips, lapses, and mistakes – a scheme that was used in several earlier studies (e.g., Reason, 1990; Wiegmann and Shappell, 1997; Sellen, 1990) – , and to supplement this approach by further analyzing mistakes in terms of the performance level at which they occur – rule-based or knowledge-based errors. The commission-omission distinction and the skill-based performance level involve the problem that one single category (errors of commission and skill-based errors) covers two very distinct error types (mistake and slip versus slip and lapse) and thus may mask important differences between them.

The Relationship Between Error Type and the Likelihood/Source of Error Detection

We expected ATC to play a critical role in the detection of most of the errors in our database, with the exception of slips which are assumed to be detected by the operator him/herself (Reason, 1990). Our expectation was confirmed (see Figure 8) – in fact, as shown earlier by Degani et al. (1991), ATC detected the majority of all types of error. This does not, of course, mean that ATC is the most efficient source of error detection. It merely reflects the fact that ASRS reports tend to be filed to gain immunity for violations that were observed by the controller. It is still interesting to see that such a considerable number of errors goes unnoticed for a long period of time and requires intervention by the last layer of defense in the system (Reason, 1990; Maurino, 1995, Woods et al., 1994). This indicates the need for better cockpit based decision support to ensure that errors, and, in particular, lapses and mistakes, are detected early on before they can lead to a potential threat or become difficult or impossible to recover from.

This finding raises a number of important issues. In the current air traffic system, pilots do not form their own intentions. Rather, their goals and targets are provided to them by the air traffic controller via clearances and requests. It is well known that numerous breakdowns occur in the communication between air and ground (Monan, 1986; Cushing, 1994) which can result in a misunderstanding about intentions. If pilots misunderstand the controller's clearance, it is impossible for them to detect their resulting erroneous actions since these actions are in accordance with the (misunderstood) clearance. And the pilot does not have enough information about the overall traffic

situation to infer that the clearance may have been misunderstood. In other words, in the current system, ATC (or possibly the other pilot who can listen to ATC communication) is the most likely source of error detection. However, this situation may change and needs to be carefully considered in the current plans for future air traffic management operations where pilots are expected to be allowed to change their flight path without permission from the ground. This means that ATC may no longer have information about pilot intent and can not, therefore, evaluate the appropriateness of pilot actions. Pilots, on the other hand, would/will have a reference - their own intentions - that they can compare their actions against. However, they may not have the information necessary to evaluate the appropriateness of their intentions given the overall traffic configuration. Thus, error detection will become more challenging, and the last layer of defense - ATC - may become much less effective.

We were surprised to find the pilots detecting many of their own knowledge-based errors (see Table 8) since the existing literature (Reason, 1990; Woods, 1987; Van Eekhout, 1981) suggests that breakdowns in on-line problem solving are the most difficult to notice. The other crewmember detected approximately the same overall percentage of errors as the pilot committing the error. However, the other crewmember was somewhat more effective in noticing errors of commission, i.e., slips and mistakes, while the pilot committing the error detected more of the lapses and errors of omission. Based on the existing literature (Reason, 1990; Sellen, 1990) we expected that the operator committing the error would fail to detect lapses. Instead, our data suggest the opposite - - the operator him/herself was quite successful in detecting their own lapses and errors of omission (see Tables 6 and 7) but they did not necessarily do so in a timely manner. Since the operator is less likely to monitor for changes and progress when he/she has not executed any action (as in the case of errors of omission) they are likely to "catch" their error only when performing a routine evaluation of the aircraft and system state(s).

This assumption regarding routine evaluations seems to be confirmed by our finding that a routine check was the most frequent source of error detection for errors of omission or lapses (50% and 47.2% respectively, see Tables 9 and 10. In other words,

while these errors are not detected immediately based on active expectation-driven information search, they are eventually caught as a result of a routine check. This explains why these errors went unnoticed for a relatively long time and resulted in some violation that required reporting. The monitoring process of the operator may in fact breakdown because the errors take familiar, and high-frequency forms such that they are "disguise(d) –by-familiarity" (Reason, 1990). For slips and mistakes, the picture is not as clear. Slips were detected almost equally often by a routine or suspicious check, by the outcome of the action unrelated to aircraft performance/behavior, or based on aircraft performance/display. If a check alone were sufficient to detect slips we would expect that this would be the most frequent detection mechanism, however, detection also depends upon the availability of cues that the action has in some way diverted (Reason, 1990). Detection of mistakes may in fact be impeded by limited information, and the tendency of individuals to accept only partial agreement between the actual state of the world and their intentions (Reason, 1990). The most frequent detection cues or mechanisms for mistakes were the outcome of an action unrelated to aircraft performance/behavior and routine checks. Detection of knowledge-based errors based on the outcome of an action unrelated to aircraft performance/behavior (see Table 11) suggests that these errors are detected once the individual is able to make a comparison between their implemented solution and the intended outcome. Our findings are thus different from those obtained in previous studies (e.g., Sellen, 1990; Allwood, 1984, or Rizzo et al., 1987) which found that slips were abruptly detected by a check, while mistakes were detected by the unexpected outcome of an action. Mistakes were detected based on routine progress checks. This difference may be explained by the fact that these studies are not comparable with the present study because they either focused on a subset of errors only (see Sellen who excluded lapses from her study) or because they involve tasks and environments that are very dissimilar from the ones in our study (see Sellen who studied everyday errors or Allwood who investigated statistical problem solving).

The field of aviation differs in various ways from other domains that were examined in earlier studies (e.g., Sellen, 1990; Allwood, 1984). Aviation is characterized by a much higher level of complexity, dynamism, and risk. Operators have to operate

highly sophisticated equipment to perform their tasks. The tasks tend to be event-driven rather than self-paced. And breakdowns in performance affects not only the pilot but potentially a large number of people on the aircraft and on the ground. Also, operators in the aviation domain are highly trained to perform their tasks. The aviation domain is highly regulated, and the pilots' actions are often determined as well as monitored by some external agent such as ATC. These domain characteristics can be expected to affect the nature of errors and error detection processes. Mistakes may be more frequent since intentions are not always formed by the operator him/herself but rather determined by some other agent. Miscommunication between the two agents can result in inappropriate goals and thus actions. The event-driven nature of aviation operations affords less pre-planning and thus tends to result in more situations involving time pressure and competing demands. This, in turn, can result in more slips and lapses due to distractions and overload. At the same time, error detection by the individual may be more common in other domains where there are fewer layers of defenses. In aviation, a large number of players monitor each other closely to avoid costly errors and their potentially disastrous consequences.

The Impact of Modern Automation Technology on Error Forms and Error Detection

Numerous authors (e.g., Woods et al., 1994) have suggested that the nature of an artifact such as modern automation technology has an impact on the nature and likelihood of errors. Since the aviation domain has seen a considerable change in terms of flight deck and aircraft technology from conventional to highly advanced glass cockpit aircraft, we were interested in exploring the impact that this technology change may have on the nature of, and reasons for, problems. One prediction was that omission errors/lapses would be more frequent on advanced aircraft in the sense that these aircraft are far more independent and can perform actions on their own. As a result, pilots may be more likely to miss undesired changes and events and fail to intervene with those activities – an error of omission (Sarter and Woods, 1995, 1997; O'Hare, 1990; Wiener; 1988). However, no significant differences between conventional and automated aircraft were found (see Figures 10, 11, and 12). This was true even when we limited our analysis to errors that

were related to flight path control - the major domain of the core of flight deck automation, the Flight Management System. The absence of the expected effect may be explained in a number of ways. It is possible that the frequency of errors of omission does, in fact, increase but that, at the same time, the changed role of the pilot from active to supervisory control supports him/her in the detection of these errors. The net effect would be that the number of omission errors that are reported to the ASRS does not increase. It is also possible that pilots on conventional aircraft simply need to perform more actions which affords a larger number of omissions. This interpretation may account for the findings shown in figure 11 which suggest a slight shift towards more lapses (as compared to slips and mistakes) on the conventional aircraft. Clearly, additional work is needed to examine these trends and possible explanations in more detail.

Another prediction related to automated versus conventional aircraft appears to be confirmed by our data. The pilot-not-flying on the automated aircraft commits relatively more errors (see Figure 13), in those cases where we could identify who was the source of the error. This can be explained by the fact that the roles and responsibilities of the pilot-flying and the pilot not-flying have changed with the addition of more automation to the flight deck. While control of the flight path on conventional aircraft is under the control of the pilot-flying, this task is shared between the two pilots on the modern flight deck where the pilot not-flying is responsible for entering some of the data (in particular, the target altitude which is the problem in many of the reported incidents) into the Flight Management System. This affords more slips or skill-based errors - as evidenced in our data (Tables 13 and 14, respectively) - but not necessarily more mistakes since the pilot-not-flying is not engaged in problem-solving activities related to the automation.

On automated flight decks, the crewmember not committing the error is more effective in detecting errors than the crewmember not committing the error on conventional flight decks (see Table 15). This is opposite to our prediction which was based on the assumption that it has become more difficult for both crewmembers to observe all activities - and thus notice erroneous actions or inputs - by their colleague (e.g., entries to the Flight Management System). However, our finding can help support

Degani et al.'s finding (1991) that, overall, flightcrews on automated aircraft detect more altitude deviations than their counterparts on conventional aircraft. This may be due to additional displays of the target altitude on automated aircraft. It is not clear whether the other crew member, in fact, detects more errors on automated flight decks or whether delayed detection by the other crewmember leads to more reports to the ASRS database.

Concluding Remarks

The results of this work highlight the necessity to better support operators in the detection of errors, in particular in the detection of lapses and mistakes. Currently, ATC serves as the last layer of defense and thus prevents many incidents from turning into accidents. Earlier detection of errors is desirable to ensure that errors can indeed be corrected before they combine with other circumstances to create a problem or even a catastrophic outcome. Also, it is not clear that ATC will be available and effective as a last layer of defense in the envisioned air traffic management system where pilots have more flexibility in choosing their flight paths without permission from the ground.

One way of better supporting error detection is through improved feedback which appears particularly important in the case of lapses or errors of omission where, given the absence of an action, the operator fails to actively search for information and the currently available feedback is not always salient enough to capture his/her attention and point out the problem. This seems to be confirmed by our finding that routine checks were most often the source of error detection. In other words, errors were detected eventually but not necessarily as soon as possible.

Another important challenge is to better support shared knowledge of intent among operators. This is suggested by our findings that ATC detected the majority of mistakes, i.e., errors in intention formation. In the current system, ATC sets goals for pilots which are often misunderstood by the crew (Monan, 1986). As a result, there is a mismatch between actual and assumed controller intent. Since error detection tends to be based on a comparison of intention and action, the pilot has no chance to detect these errors - - his/her actions are in accordance with the assumed controller intentions. Only ATC knows that the observed aircraft behavior does not match the given clearance. One

possible way of improving the situation may be the introduction of digital communication which will allow for the uplink of controller clearances to the flight deck (Wickens et al., 1997). Pilots may still misread the displayed or printed messages; however, the clearance is available for later reference and may even be available to the aircraft automation which could compare clearance and aircraft behavior and indicate discrepancies to the pilot.

Another important step that is suggested by our data is to minimize factors that can lead to inattention (see Table 5). This may be achieved by means of improved task and resource management to minimize competing demands and by even stricter "sterile cockpit" policies to avoid distractions by flight attendants or cockpit observers.

Finally, it seems important to investigate in more detail the impact of automation technology on the nature and detection of errors. Numerous authors have proposed that the design of an artifact shapes the form and likelihood of error (e.g., Woods et al., 1994; Reason, 1990). This assertion was only partially supported by our data (see Tables 12-14 and Figure 13) which suggest that the new role of the pilot-not-flying on the automated flight deck affords more opportunities for committing errors. This appeared to be counterbalanced, however, by the observed increased likelihood of error detection by the pilot not committing the error. It is possible that the use of different error classification schemes or a more in-depth process analysis of incidents will reveal additional differences between erroneous actions and assessments on conventional versus automated aircraft. Awareness and a better understanding of those differences is critical given our goal is to reduce the accident rate in the future air traffic environment which will most likely be dominated by advanced technology aircraft.

Finally, we would like to emphasize the need for collecting more systematic data on error detection mechanisms and failures. Despite the importance of supporting error management, very little research has been conducted in this area, and data from operational environments are limited. In particular, we think that the ASRS database could provide important insights into error detection. However, currently, reporters are not encouraged to provide detailed information about the processes leading to the detection of an error.

Appendix

A. ASRS Reporting Form

DO NOT REPORT AIRCRAFT ACCIDENTS AND CRIMINAL ACTIVITIES ON THIS FORM.
ACCIDENTS AND CRIMINAL ACTIVITIES ARE NOT INCLUDED IN THE ASRS PROGRAM AND SHOULD NOT BE SUBMITTED TO NASA.
ALL IDENTITIES CONTAINED IN THIS REPORT WILL BE REMOVED TO ASSURE COMPLETE REPORTER ANONYMITY.

(SPACE BELOW RESERVED FOR ASRS DATE/TIME STAMP)

IDENTIFICATION STRIP: *Please fill in all blanks to ensure return of strip.*
NO RECORD WILL BE KEPT OF YOUR IDENTITY. This section will be returned to you.

TELEPHONE NUMBERS where we may reach you for further details of this occurrence:

HOME Area ______ No. ________ - __________ Hours _____
WORK Area ______ No. ________ - __________ Hours _____

NAME ______________________________
ADDRESS/PO BOX ______________________________
____
CITY ______________ STATE ____ ZIP ________

TYPE OF EVENT/SITUATION ______________
____
DATE OF OCCURRENCE ______________
LOCAL TIME (24 hr. clock) ______________

PLEASE FILL IN APPROPRIATE SPACES AND CHECK ALL ITEMS WHICH APPLY TO THIS EVENT OR SITUATION.

REPORTER	FLYING TIME	CERTIFICATES/RATINGS		ATC EXPERIENCE
o Captain o First Officer o pilot flying o pilot not flying o Other Crewmember o ________	total ________ hrs. last 90 days ______ hrs. time in type ______ hrs.	o student o commercial o instrument o multiengine o ________	o private o ATP o CFI o F/E	o FPL o Developmental radar ________ yrs. non-radar ________ yrs. supervisory ________ yrs. military ________ yrs.

AIRSPACE		WEATHER		LIGHT/VISIBILITY		ATC/ADVISORY SERV.	
o Class A (PCA) o Class B (TCA) o Class C (ARSA) o Class D (Control Zone/ATA) o Class E (General Controlled) o Class G (Uncontrolled)	o Special Use Airspace o airway/route ______ o unknown/other ______ ________ ________ ________	o VMC o IMC o mixed o marginal o rain o fog	o ice o snow o turbulence o tstorm o windshear o ______	o daylight o dawn ceiling ______ visibility ______ RVR ______	o night o dusk feet miles feet	o local o ground o apch o dep	o center o FSS o UNICOM o CTAF
						Name of ATC Facility: ________	

	AIRCRAFT 1			AIRCRAFT 2		
Type of Aircraft (Make/Model)	(Your Aircraft) ________		o EFIS o FMS/FMC	(Other Aircraft) ________		o EFIS o FMS/FMC
Operator	o air carrier o commuter	o military o private	o corporate o other ______	o air carrier o commuter	o military o private	o corporate o other ______
Mission	o passenger o cargo	o training o pleasure	o business o unk/other ______	o passenger o cargo	o training o pleasure	o business o unk/other ______
Flight plan	o VFR o IFR	o SVFR o DVFR	o none o unknown	o VFR o IFR	o SVFR o DVFR	o none o unknown
Flight phases at time of occurrence	o taxi o takeoff o climb	o cruise o descent o approach	o landing o missed apch/GAR o other ______	o taxi o takeoff o climb	o cruise o descent o approach	o landing o missed apch/GAR o other ______
Control status	o visual apch o controlled o no radio	o on vector o none o radar advisories	o on SID/STAR o unknown	o visual apch o controlled o no radio	o on vector o none o radar advisories	o on SID/STAR o unknown

If more than two aircraft were involved, please describe the additional aircraft in the "Describe Event/Situation" section.

LOCATION	CONFLICTS
Altitude ________________ o MSL o AGL	Estimated miss distance in feet: horiz ______ vert ______
Distance and radial from airport, NAVAID, or other fix ________	Was evasive action taken? o Yes o No
________________	Was TCAS a factor? o TA o RA o No
Nearest City/State ________________	Did GPWS activate? o Yes o No

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

AVIATION SAFETY REPORTING SYSTEM

NASA has established an Aviation Safety Reporting System (ASRS) to identify issues in the aviation system which need to be addressed. The program of which this system is a part is described in detail in FAA Advisory Circular 00-46C. Your assistance in informing us about such issues is essential to the success of the program. Please fill out this form as completely as possible, enclose in an sealed envelope, affix proper postage, and and send it directly to us.

The information you provide on the identity strip will be used only if NASA determines that it is necessary to contact you for further information. THIS IDENTITY STRIP WILL BE RETURNED DIRECTLY TO YOU. The return of the identity strip assures your anonymity.

Section 91.25 of the Federal Aviation Regulations (14 CFR 91.25) prohibits reports filed with NASA from being used for FAA enforcement purposes. This report will not be made available to the FAA for civil penalty or certificate actions for violations of the Federal Air Regulations. Your identity strip, stamped by NASA, is proof that you have submitted a report to the Aviation Safety Reporting System. We can only return the strip to you, however, if you have provided a mailing address. Equally important, we can often obtain additional useful information if our safety analysts can talk with you directly by telephone. For this reason, we have requested telephone numbers where we may reach you.

Thank you for your contribution to aviation safety.

NOTE: *AIRCRAFT ACCIDENTS SHOULD NOT BE REPORTED ON THIS FORM. SUCH EVENTS SHOULD BE FILED WITH THE NATIONAL TRANSPORTATION SAFETY BOARD AS REQUIRED BY NTSB Regulation 830.5 (49CFR830.5).*

Please fold both pages (and additional pages if required), enclose in a sealed, stamped envelope, and mail to:

NASA AVIATION SAFETY REPORTING SYSTEM
POST OFFICE BOX 189
MOFFETT FIELD, CALIFORNIA 94035-0189

DESCRIBE EVENT/SITUATION

Keeping in mind the topics shown below, discuss those which you feel are relevant and anything else you think is important. Include what you believe really caused the problem, and what can be done to prevent a recurrence, or correct the situation. (USE ADDITIONAL PAPER IF NEEDED)

Page 2 of 2

CHAIN OF EVENTS
- How the problem arose
- How it was discovered
- Contributing factors
- Corrective actions

HUMAN PERFORMANCE CONSIDERATIONS
- Perceptions, judgments, decisions
- Actions or inactions
- Factors affecting the quality of human performance

B. Data Analysis Form

ASRS Data Collection

Incident Number:____________________

Aircraft Type: ______________________

___ Conventional
___ Automated

Which crew member committed the error?

___ PF
___ PNF
___ Other

___ Captain
___ First Officer

Which crew member detected the error?

___ PF
___ PNF
___ Other

___ Captain
___ First Officer

Short Summary of the Incident:

____
____
____
____
____

Error Phenotype:

___ Altitude Deviation (how much: ________________)
___ Heading/Course Deviation (how much: __________)
___ Speed Deviation (how much: __________________)
___ Runway Incursion
___ Other

Error Classification:

___ Omission (fails to take required action)
___ Commission (performs inappropriate action or performs action inappropriately)

(con't)

___ Slip (performs intended action inappropriately)
___ Lapse (forgets to take intended action)
___ Mistake (deficiency in intention formation, or means to achieve goal)

Performance Level:

___ Skill-based performance (routine task – highly practiced)
___ Rule-based performance (solving a problem for which a solution/rule exists/is known)
___ Knowledge-based performance (encountering a novel problem/situation – on-line problem-solving by trial and error)

Contributing Factors:

___ Lack of knowledge/understanding
___ Inattention
___ Distraction
___ Time Pressure
___ Competing Demands/High Workload

Detection of Error (indicate the sequence if more than one applies):

Who detected the error?

___ Operator who committed the error
___ Other crewmember
___ Air Traffic Control
___ Other Ground Personnel
___ Other

Detection Cue/Mechanisms:

___ Outcome of an action (other than aircraft performance)
___ Routine check
___ Suspicious check
___ Limiting function
___ Alarm
___ Aircraft performance/displays
___ Other

Bibliography

Allwood, C. (1984). Error detection processes in statistical problem solving. *Cognitive Science*, **8**, 413-437.

Bagnara, S., Stablum, F., Rizzo, A., Fontana, A., and Ruo, M. (1987). Error detection and correction: A study on human-computer interaction in a hot strip mill production planning and control system. *In Preprints of the First European Meeting on Cognitive Science Approaches to Process Contol.* Marcoussis, France, October.

Barker, K. (1962). Detecting errors in hospitals. *American Journal of Hospital Pharmacy*, **19**, 361-369.

Beaty, D. (1995). *The Naked Pilot.* Shrewsbury, England: Airlife Publishing Limited.

Chappell, S. (1994). Using voluntary incident reports for human factors evaluations. In N. Johnson, N. McDonald, and R. Fuller (Eds.), *Aviation Psychology in Practice.* Brookfield, VT: Ashgate Publishing Company.

Chou, C., Madhavan, D., and Funk, K. (1996). Studies of cockpit task management errors. *The International Journal of Aviation Psychology,* **6** (4), 307-320.

Coles, M., Scheffers, M., and Fournier, M. (1995). "Where did you go wrong?" Errors Partial errors, and the nature of human information processing. *Acta Psychologica*, **90** (1-3), 129-144.

Cushing, S. (1994). *Fatal Words: Communication Clashes and Aircraft Crashes.* Chicago, IL: The University of Chicago Press.

Degani, A., Chappell, S., and Hayes, M. (1991). Who or what saved the day? A comparison of traditional and glass cockpits. In *Proceedings of Sixth International Symposium on Aviation Psychology.* Columbus, OH.

Diehl, A. (1991). Human performance and systems safety considerations in aviation mishaps. *The International Journal of Aviation Psychology*, **1** (2), 97-106.

Farmer, E. (1988). Cognitive processes. In J. Ernstein and P. King (Eds.). *Aviation Medicine, 2nd Edition.* Linacre House, Jordan Hill, Oxford: Butterworth-Heinemann Ltd.

Fitts, P., and Jones, R. (1947). Analysis of factors contributing to 460 "pilot-error" experiences in operating aircraft controls. In H. W. Sinaiko (Ed.). *Selected Papers on Human Factors in the Design and Use of Control Systems.* New York, NY: Dover.

Harle, P. (1994). Investigation of human factors: The link to accident prevention. In N. Johnson, N. McDonald, and R. Fuller (Eds.), *Aviation Psychology in Practice*. Brookfield, VT: Ashgate Publishing Company.

Heinrich, H., Peterson, D., and Roos, N. (1980*). Industrial Accident Prevention, 5th Edition*. New York, NY: McGraw-Hill Inc.

Hollnagel, E. (1993). The phenotype of erroneous actions. *International Journal of Man-Machine Studies*, **39**, 1-32.

Kroll, M., and Ford, M. (1992). The illusion of knowing, error detection, and motivational orientations. *Contemporary Eductional Psychology*, **17**, 371-378.

Loftus, E. (1979). *Eyewitness Testimony*. Cambridge, MA: Harvard University Press.

MacKay, D. (1992). Awareness and error detection: New theories and research paradigms. *Consciousness and Cognition,* **1**, 199-225.

Maurino, D., Reason, J., Johnson, N., and Lee, R. (1995*). Beyond Aviation Human Factors*. Brookfield, VT: Ashgate Publishing.

Monan, W. (1986). Human factors in aviation operations: The hearback problem. *NASA Contractor Report #177398.*

Norman, D. A. (1981). Categorization of action slips. *Psychological Review*, **88** (1), 1-15.

O'Hare, D., and Roscoe, S. (1990*). Flightdeck Performance: The Human Factor*. Ames, IO: Iowa State University Press.

O'Hare, D., Wiggins, M., Batt, R., and Morrison, D. (1994). Cognitive failure analysis for aircraft accident investigation. *Ergonomics*, **37** (11), 1855-1869.

Orasanu, J., and Fischer, U. (1997). Finding decisions in natural environments: The view from the cockpit. In C. Zsambok and G. Klein (Eds.), *Naturalistic Decision Making*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Rabbitt, P., and Phillips, S. (1967). Error-detection and correction latencies as a function of S-R Compatibility. *Quarterly Journal of Experimental Psychology,* **19**, 37-42.

Rabbitt, P. (1978). Detection of errors by skilled typists. *Ergonomics*, **21** (11), 945-958.

Rabbitt, P., Cumming, G., and Vyas, S. (1978). Some errors of perceptual analysis in visual search can be detected and corrected. *Quarterly Journal of Experimental Psychology,* **30**, 319-332,

Rasmussen, J. and Jensen, A. (1974). Mental procedure in real-life tasks: A case study of electronic trouble shooting. *Ergonomics*, **17** (3), 293-307.

Reason, J. (1990). *Human Error.* New York, NY: Cambridge University Press.

Reason, J. (1995). A systems approach to organizational error. *Ergonomics,* **38** (8), 1708-1721.

Rizzo, A., Bargnara, S., and Visciola, M. (1987). Human error detection processes. *International Journal of Man-Machine Studies,* **27**, 555-570.

Rouse, W. and Morris, L. (1987). Conceptual design of a human error tolerant interface for complex engineering systems. *Automatica*, **23** (2), 231-235.

Sarter, N. B., and Woods, D. D. (1995). Strong, silent, and out-of-the-loop: Properties of advanced (cockpit) automation and their impact on human-automation interaction. *Cognitive Systems Engineering Laboratory Technical Report CSEL-TR-01*, February.

Sarter, N. B., Woods, D.D., and Billings, C. E. (1997). Automation surprises. In G. Salvendy (Ed*.), Handbook of human factors and ergonomics (2nd ed)*. New York, NY: Wiley-Interscience Publication.

Schiavo, M. F. (1998). At high risk. *The American Legion Magazine*, February.

Sellen, A. (1990). Mechanisms of human error and human error detection. Unpublished Ph.D. thesis, University of California, San Diego.

Scheffers, M., Coles, M. Berstein, Gehring, W., and et al. (1996). Event-related brain potentials and error-related processing: An analysis of incorrect responses to go and no-go stimuli. *Psycholphysiology,* **33** (1), 42-53.

Smith, H. P. (1979). A simulator study of the interaction of pilot workload with errors, vigilance, and decisions. *NASA Technical Memorandum #78482*.

Thompson, D. (1980). Commercial air crew detection of systems failures: State of the art and future trends. In J. Rasmussen and W. Rouse (Eds.). *Human Detection and Diagnosis of System Failures*. New York, NY: Plenum Press.

Van Eekhout, J., and Rouse, W. (1981). Human errors in detection, diagnosis, and compensation for failures in the engine control room of a supertanker. *IEEE Transactions on Systems, Man, and Cybernetics,* **11** (12), 813-816.

Weiner, E. (1988). Cockpit automation. In E. Wiener and D. Nagel (Eds*.). Human Factors in Aviation*. San Diego, CA: Academic Press.

Wickens, C. D., and Flach, J. M. (1988). Information processing. In E. L. Wiener and D. C. Nagel (Eds.), *Human Factors in Aviation.* San Diego, CA: Academic Press.

Wickens, C.D. (1993). Timesharing, workload, and human error. In K. Cardosi, and M. S. Huntley (Eds.), *Human Factors for Flight Deck Certification Personnel.* Springfield, VA: National Technical Information Service.

Wickens, C. D., and McCloy, T. (1993). ASRS and Aviation Psychology. In *Proceedings of the Seventh International Symposium on Aviation Psychology.* Columbus, OH.

Wickens, C. D. (1995). Aerospace techniques. In J. Weimer (Ed.), *Research Techniques in Human Engineering.* Englewood Cliffs, NJ: Prentice Hall.

Wickens, C. D., Mayor, A. S., and McGee, J. P. (1997). *Flight to the Future: Human Factors in Air Traffic Control.* Washington, D.C.: National Academy Press.

Weigmann, D. A, and Shappell, S. A. (1997). Human factors analysis of postaccident data: Applying theoretical taxonomies of human error. *The International Journal of Aviation Psychology,* 7 (1), 67-81.

Woods, D., O'Brien, J., and Hanes, L. (1987). Human factors challenges in process control: The case of nuclear power plants. In G. Salvendy (Ed.), *Handbook of Human Factors.* New York, NY: Wiley-Interscience Publication.

Woods, D., Johannesen, L., Cook, R., and Sarter, N. (1994). *Behind Human Error.* Wright-Patterson AFB, OH: CSERIAC.